(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
11 August 2005 (11.08.2005)

PCT

(10) International Publication Number
WO 2005/072262 A2

(51) International Patent Classification: Not classified

(21) International Application Number:
PCT/US2005/001983

(22) International Filing Date: 21 January 2005 (21.01.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/763,712 21 January 2004 (21.01.2004) US

(71) Applicant *(for all designated States except US)*: **SOLAZYME, INC.** [US/US]; Edison Technology Park, 3475-T Edison Way, Menlo Park, CA 94025 (US).

(72) Inventor: **DILLON, Harrison, F.**; Solazyme, Inc., Edison Technology Park, 3475-T Edison Way, Menlo Park, CA 94025 (US).

(74) Agent: **DILLON, Harrison, F.**; Solazyme, Inc., Edison Technology Park, 3475-T Edison Way, Menlo Park, CA 94025 (US).

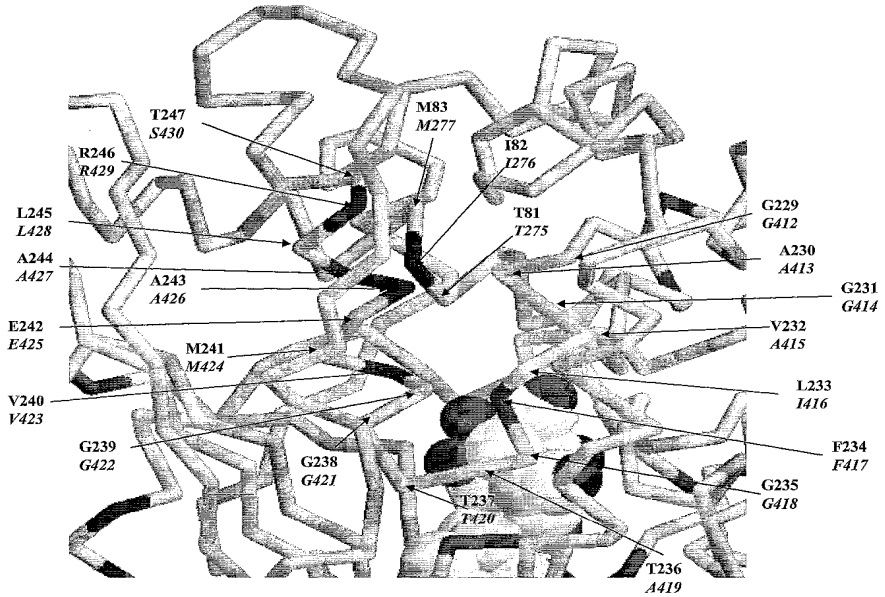
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHODS AND COMPOSITIONS FOR EVOLVING MICROBIAL HYDROGEN PRODUCTION



(57) Abstract: The invention provides methods and compositions for engineering cells to generate large amounts of hydrogen. Genes that are involved in hydrogen production pathways and genes that are upregulated when cells are exposed to conditions conducive to the generation of hydrogen are mutagenized according to disclosed protocols. Microbes containing nucleic acid constructs are screened or selected for the ability to generate an increased amount of hydrogen. Methods of producing hydrogen are also disclosed.

WO 2005/072262 A2

Methods and compositions for evolving microbial hydrogen production

[001] This application claims priority to U.S. Patent Application No.: 10/287,750, filed November 4, 2002. This application also claims priority to U.S. Patent Application No.: 10/411,910, filed April 12, 2003. This application also claims priority to U.S. Patent Application No.: 60/500,032, filed September 3, 2003. U.S. Patent Applications 10/287,750, 10/411,910, and 60/500,032 are hereby fully incorporated by reference for all purposes.

BACKGROUND OF THE INVENTION

[002] Hydrogen is the most abundant element on earth. When hydrogen is burned as a fuel, the only byproducts are heat and water. Large-scale commercial production of hydrogen could have a massive impact on the world environment and economy. The availability of an environmentally clean, renewable energy source would greatly curtail if not end large-scale dependence on fossil fuels. Hydrogen can be converted into electrical energy by utilizing fuel cells, but it would also be an ideal replacement for oil-based energy since it has a calorie per unit weight of 3 to 4 times that of petroleum (United States Patent 4,532,210).

[003] Fuel cell technology is being developed at a rapid pace, however a plentiful and commercially viable source of hydrogen with which to run fuel cells has not yet been created. There are a variety of known methods for producing hydrogen. For instance, inorganic membrane electrolysis technology (IMET) involves the splitting of water through electrolysis in the reaction $2H_2O \Rightarrow 2H_2 + O_2$. Water electrolysis occurs through passing an electric current through water to separate it into hydrogen and oxygen. Hydrogen gas is produced at the negative cathode and oxygen gas is produced at the positive anode. Another source of hydrogen production is through reforming natural gas. Unfortunately this process produces carbon dioxide making this source of hydrogen less than ideal.

[004] Hydrogen production through electrolysis, powered by renewable sources such as wind, solar energy through photovoltaic cells, or hydroelectric power has the advantage of not creating pollutants in the process of generating hydrogen, however the potential amount of hydrogen that can be produced through these methods may be limiting.

[005] What is needed are methods for engineering microbial organisms to produce hydrogen for extended periods of time in large amounts, something no known microbe is currently capable of doing. Furthermore, methods of identifying genes that are involved in hydrogen production pathways of microbes so that they can be optimized for efficient contribution to the production of hydrogen are needed.

BRIEF SUMMARY OF THE INVENTION

[006] Provided are sethod for engineering a cell to produce an increased amount of hydrogen comprising providing a mutagenized nucleic acid sequence derived from a first gene that encodes a protein involved in a hydrogen production pathway, transforming a cell with the mutagenized nucleic acid sequence, and screening or selecting the cell for an increased amount of hydrogen.

[007] Methods are provided for identifying a first independent transformant which produces an increased amount of hydrogen, recovering the mutagenized nucleic acid sequence from the independent transformant, further mutagenizing the recovered mutagenized nucleic acid sequence to create a new library of mutagenized nucleic acid sequences, transforming cells with the new library of mutagenized nucleic acid sequences, and screening or selecting for a new independent transformant that generates an increased amount of hydrogen compared to the first independent transformant.

[008] In some methods a plurality of mutagenized nucleic acid sequences are recovered from a plurality of independent transformants which produce an increased amount of hydrogen, wherein the plurality of mutagenized nucleic acid sequences are subjected to gene reassembly to generate the new library.

[009] In one embodiment a plurality of mutagenized nucleic acid sequences are used to transform a population of cells, followed by the screening or selecting.

In one embodiment the first gene is selected from the group that encodes ferredoxin, catalase, isoamylase, malate dehydrogenase, 14-3-3 protein, enolase, aldolase, ribosomal protein S8, ribosomal protein L17, ribosomal protein S18, ribosomal protein L37, ribosomal protein L12, ribosomal protein S15, iron-hydrogenase, nickel-iron hydrogenase, and components of the photosystem I, photosystem II, light harvesting antenna and cytochrome b_6-f complexes.

[010] The methods provided include mutagenesis of iron hydrogenase proteins including mutagenesis of the $X^1X^2X^3FX^4X^5X^6GGVMEAAX^7R$ and ADX^8TIX^9EE segments. In some methods, cognate sequences of these conserved segments of iron hydrogenases are substituted into a Chlamydomonas iron hydrogenase. In some methods, gene reassembly methods are performed in which a Chlamydomonas iron hydrogenase is mutagenized by incorporation of segments of iron hydrogenase proteins from other species. Preferred segments for inclusion in gene reassembly include segments that form parts of the gas channel, also referred to as the gas channel. In some methods a higher molecular weight amino acis is substituted into a gas channel segment, such as a tryptophan for the methionine in the C. reinhardtii TIMEE segment. In other gene reassembly methods the iron hydrogenase is reassembled using methods that involve attaching sections of duplex DNA that have only one overhanging nucleotide. In other methods oligonucleotides encoding gas channel segments are annealed to a scaffold nucleic acid, where the oligonucleotides anneal to non-overlapping sites. Preferably, the mutagenesis of a hydrogenase does not decrease the protein's ability to accept electrons from an electron donor. In some methods the mutagenized nucleic acid is transcribed by a light-driven promoter.

[011] Methods are provided herein for screening or selecting for a hydrogen production phenotype in the presence of oxygen at a concentration selected from the ranges comprising more than 0.5%, more than 5.0%, more than 10%, more than 15%, approximately 21%, more than 21%, more than 25%, more than 30% or more than 35% oxygen. In some methods the cells screened or selected are in liquid culture media.

[012] Methods are provided for mating (a) at least one cell of a strain containing a mutagenized form of the first gene, wherein the at least one cell is identified by the screening or selecting or wherein the at least one cell is derived through mating from a cell identified by the screening or selecting; (b) to at least one cell of a distinct strain containing a mutagenized form of the second gene, wherein the at least one cell is identified by the screening or selecting, or wherein the at least one cell is derived through mating from a cell identified by the screening or selecting; and (c) screening or selecting for a progeny cell that produces an increased amount of hydrogen compared to any parent cell.

[013] A method of hydrogen production is disclosed, comprising placing cell containing a mutagenized nucleic acid sequence corresponding to a gene that is involved in a hydrogen production pathway into liquid culture media or on to solid culture media, wherein the mutagenized nucleic acid sequence is operably linked to a transcriptional promoter sequence; culturing said transformed cell under conditions sufficient to stimulate transcription of said mutagenized nucleic acid sequence(s); and collecting an evolved gas. In some methods the culture media supplied to the cells is photoautotrophic growth requiring media.

[014] Mating methods are provided. One method is a method of multiparental mating of microbes that mate in response to a stimulus, comprising: (a) providing a cell from each of 3 or more strains of microbes capable of mating to each other in culture medium; (b) providing the stimulus; (c) allowing cells to mate and produce progeny; (d) allowing the progeny cells to achieve sexual reproduction capability; (e) providing the stimulus at least one

more time; and (f) screening or selecting the further progeny for a desired phenotype. In some methods the microbes are green algae and the stimulus is the removal of nitrogen from the media and illumination by light comprising a wavelength of light between about 0.42-0.52 micrometers.

In some methods the green algae are of the Chlamydomonas genus, optionally of a species selected from the group comprising reinhardtii, eugametos, incerta, and moewusii. In other methods the stimulus is interruption of exponential growth in continuous light with a reduction in light, followed by addition of light, wherein the reduction in light occurs for a period selected from the group consisting of at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or more than 12 hours. In other methods the microbes are of the Scendesmus genus and the stimulus is the addition of chromium to the culture media. In some methods the desired phenotype is hydrogen production. In still other methods, nucleic acid exchange occurs between only two parental cells at a time during the mating process.

[015] The foregoing description of some preferred embodiments of the invention is not a limiting description of the invention, and many other embodiments of the invention are described herein.

BRIEF DESCRIPTION OF THE DRAWINGS

[016] Figure 1 demonstrates the method of subjecting homologous genes cloned from different microbes capable of producing hydrogen to Dnase I digestion in preparation for DNA shuffling procedures.

[017] Figure 2 demonstrates the construction of a library of shuffled sequences. Dnase I digested fragments are annealed to chimeric oligonucleotides that contain sequences corresponding to the N and C terminal ends of the coding regions of the shuffled genes as well as linker sequences referred to as "unique sequences" that are present at both ends of each fragment after annealing and primerless PCR.

[018] Figure 3 demonstrates the denaturation, annealing, and primerless PCR of DNA fragments containing different elements of a DNA construct used to transform cells. Denatured fragments anneal through unique sequences to other fragments. The shuffled library of coding regions of shuffled differentially regulated genes is flanked by unique sequences that anneal to promoter and transcriptional terminator sequences.

[019] Figure 4 depicts a map of the DNA constructs described in Example 1, with details demonstrating the annealing points of each shuffled library to flanking nonshuffled segments during construction.

[020] Figure 5 depicts a map of the DNA constructs described in Example 1.

[021] Figure 6 depicts a detailed map of the DNA constructs described in Example 1, including the relative positions of PCR primers and chimeric oligonucleotides. The map is not necessarily drawn to scale.

[022] Figure 7 depicts a detailed map of the DNA constructs described in Example 2, including the relative positions of PCR primers and chimeric oligonucleotides. The map is not necessarily drawn to scale.

[023] Figure 8 depicts a screening system for use with liquid culture-containing multiwell plates.

[024] Figure 9 depicts amino acid residues in and near the gas channel of the Clostridium pasteurianum iron hydrogenase from the structure 1feh in the Protein Data Bank. The amino acid positions from the Clostridium pasteurianum iron hydrogenase are shown in italics, while the corresponding amino acid positions from a Chlamydomonas reinhardtii iron hydrogenase are shown above in non-italicized font, both according to the numbering from Figure 4 of Happe, Eur J Biochem (2002) Feb;269(3):1022-32.

[025] Figure 10 depicts the codon usage table of C. reinhardtii. Most preferred codons are shown underlined and in bold-face type. Any cDNA sequence can be recoded for maximal expression in C. reinhardtii by substituting non-preffered codons for most preferred codons. Codon usage tables for microbes can be found at

http://www.kazusa.or.jp/codon/.

[026] Figure 11 depicts the mating of two C. reinhardtii cells. Genetic alterations on cognate chromosomes that each increase hydrogen production can cosegregate in a progeny cell through a recombination event. Such progeny can produce more hydrogen than parental strains.

[027] Figure 12 depicts multiparental mating of four strains of C. reinhardtii. Each of the four strains has a genetic alteration that increases hydrogen production. The multiparental mating reaction proceeds through at least two cycles of nitrogen deprivation and germination. All four genetic alterations can cosegregate in a progeny cell. Such progeny can produce more hydrogen than either parent strain in any of the matings that occur in the multiparental mating reaction.

[028] Figures 13-14 depict a gene reassembly protocol for incorporating segments of diverse Iron hydrogenaserogenases into the overall framework of a single Iron hydrogenaserogenase. In this example, a C. reinhardtii Iron hydrogenaserogenase gene provides the single stranded framework. The design of the protocol allows framework/hinge regions to be retained while architecture of the gas channel is altered compared to the C. reinhardtii Iron hydrogenaserogenase.

[029] Figure 15 shows the key to the identity of the amino acids of step 1 of figure 13 and the corresponding identity of codons in nucleic acids in steps 2-9 of figures 13-14.

[030] Figure 16 shows the divergent sequences from SEQ ID Nos: 1-112 that correspond to the segments of Iron hydrogenaserogenases that line the gas channel. These are the segments that are schematically depicted in figure 13, step 1. The sequences are used to design the oligonucleotides in step 2 of figure 13.

[031] Figure 17 shows one example of how gas channel segments from SEQ ID Nos: 1-112 are reverse translated into recoded nucleotide sequence. C. reinhardtii flanking sequence is added to each side of the oligonucleotide sequence to ensure adequate annealing. Although step 1 of figure 13 depicts 3 segments, which figure 16 shows only 2 segments, the $X^1X^2X^3FX^4X^5X^6GGVMEAAX^7R$ segment is broken into two distinct segments to allow greater combinatorial diversity af the library, as this figure shows.

DETAILED DESCRIPTION OF THE INVENTION

[032] All publications, patents, patent applications, and other references cited are fully incorporated by reference for all purposes.

[033] Definitions: The following definitions are intended to convey the intended meaning of terms used throughout the specification and claims, however they are not limiting in the sense that minor or trivial differences fall within their scope.

[034] "Differential expression profile" means information about the activity of at least one gene or the presence or activity of at least one protein in a cell when the cell is exposed to at least two different environmental conditions or chemical environments. Literally any difference in the conditions that the cell might be exposed to can cause a difference in the expression of one or more genes or the presence or activity of one or more proteins.

[035] "Conditions more conducive to the generation of hydrogen" means any set of conditions under which a cell generates hydrogen.

[036] "Conditions more conducive to the generation of hydrogen" also means, in an experiment intended to generate a differential expression profile, conditions under which a cell that already generates a measurable amount

of hydrogen under a first set of conditions generates, under a second set of conditions distinct from the first set, a measurably greater amount of hydrogen than it does under the first set of conditions.

[037] "Conditions less conducive to the generation of hydrogen" means any set of conditions under which a cell either generates no measurable amount of hydrogen or generates measurably less hydrogen than under conditions more conducive to the generation of hydrogen. Specifically, conditions more conducive to the generation of hydrogen cause a cell to generate a measurable amount of hydrogen while conditions less conducive to the generation of hydrogen cause a cell to generate either no hydrogen or measurably less hydrogen than the conditions more conducive to the generation of hydrogen in that same experiment. When cells are cultured under conditions less conducive to the generation of hydrogen yet produce a measurable amount of hydrogen, that measurable amount of hydrogen is less than the amount of hydrogen produced by cells cultured under conditions more conducive to the generation of hydrogen in order to produce a differential expression profile. In terms of measuring the amount of hydrogen produced, a greater amount of hydrogen produced by a cell under one condition compared to another condition is determined by measuring production of hydrogen over a given time interval.

[038] "Conditions not conducive to the generation of hydrogen" means any set of conditions under which a cell does not generate a measurable amount of hydrogen.

[039] "Culture conditions" and "conditions" means the plurality of variables that are manipulated when culturing microbes, including but not limited to exposure to light or certain wavelengths of light, exposure to certain molecules, nutrients, elements, and the like in culture media as well as exposure to different concentrations of these molecules, elements, nutrients, and the like, temperature, placement in darkness or partial darkness, exposure to other microbes or viruses, as well as any other variable that is manipulated when culturing microbes.

[040] "Differentially regulated" means where the activity of a gene or a protein in a cell is in some way different under one set of culture conditions than under a different set of culture conditions. For instance, Chlamydomonas cells express certain genes in higher amounts during the first hour of anaerobic culturing in the dark as compared to culturing in the presence of oxygen and illumination. Even though certain genes are expressed in both culture conditions, if the genes are expressed at different levels between the two conditions they are differentially regulated.

[041] "Mutagenized nucleic acid sequence" means a nucleic acid sequence in which the nucleotide sequence of the mutagenized nucleic acid sequence differs from a starting sequence prior to mutagenesis by at least one base pair. For instance, a single nucleic acid sequence is amplified using error-prone PCR to generate a library of nucleic acid sequences that are similar in sequence to the starting sequence but differ by at least one base pair, and are therefore mutagenized nucleic acid sequences. Alternatively, a plurality of nucleic acid sequences that have significant sequence identity are put through a gene reassembly process to generate mutagenized nucleic acid sequences. Mutagenized nucleic acid sequences are derived from the full or partial sequence of at least one wild type sequence, also referred to as a starting sequence. In gene reassembly processes the starting sequences are the parental genes in non-recombined form. Mutagenized nucleic acid sequences can also be generated by chemical mutagenesis of living cells using carcinogens such as nitrosoguanidine (NTG).

[042] "Significant sequence identity" means at least 40%, preferably 50%, more preferably 60% and more preferably 70%, and even more preferably 80% or 90% or higher nucleotide sequence identity when compared using a standard sequence comparison such as the BLAST program available at www.ncbi.nlm.nih.gov. utagenized nucleic acid sequences can also be generated using standard site-directed mutagenesis protocols (Maniatis et al. (1989) Molecular Cloning : A Laboratory Manual Cold Spring Harbor Laboratory).

[043] "Downregulated" means, when relating to a gene, when a gene is transcribed less per unit time or when a

gene's corresponding RNA is translated less times per unit time than it was when compared to the level of transcription or translation previously. "Downregulated" means, when relating to a protein, when the protein's activity per unit time is diminished when compared to the level of activity per unit time previously, when the protein is degraded at a faster rate, or when the gene encoding the protein is transcribed less per unit time or is translated less times per unit time than it was when compared to the level of transcription or translation previously.

[044] "Upregulated" means, when relating to a gene, when a gene is transcribed or when a gene's corresponding RNA is translated more times per unit time than it was when compared to the level of transcription or translation previously. "Upregulated" means, when relating to a protein, when the protein's activity per unit time is increased when compared to the level of activity per unit time previously, when a protein is degraded at a slower rate, or when the gene encoding the protein is transcribed more per unit time or is translated more times per unit time than it was when compared to the level of transcription or translation previously.

[045] "Shuffling" means recombining a first nucleic acid with at least one other nucleic acid distinct in sequence from the first nucleic acid, wherein the first nucleic acid and the at least one other nucleic acid recombine through sequence-specific annealing with each other or to a third nucleic acid. Shuffling is also referred to as gene reassembly.

[046] "Site-directed mutagenesis" means generating a desired gene sequence that differs from the sequence of a starting gene, wherein the sequence difference is a specifically designed amino acid insertion, deletion, substitution, or combination thereof.

[047] "Increased amount of hydrogen" means an amount of hydrogen produced by a strain that has been transformed with a mutagenized nucleic acid sequence that is greater than the amount of hydrogen produced by the starting strain that has either not been transformed with the mutagenized nucleic acid sequence or that has been transformed using only control or vector sequences.

[048] A cell "derived through mating" from a distinct cell is a cell that would not exist but for the mating of the distinct cell with at least one other cell. For example, a distinct cell has a mutagenized nucleic acid sequence that causes increased hydrogen production. The distinct cell is mated to another cell, resulting in progeny cells. The progeny cells are derived through mating from the first cell.

DESCRIPTION

Culturing bacteria under conditions more conducive to the generation of hydrogen

[049] Methods for culturing photosynthetic bacteria under conditions more conducive and less conducive to the generation of hydrogen are known (Maness, (2001) Appl Microbiol Biotechnol Dec;57(5-6):751-6; Weaver PF, Proceedings of the Fifth Joint US/USSR Conference of the Microbial Enzyme Reactions Project, Jurmala, Latvia, USSR (1979) 461-479). Methods for culturing cyanobacteria under conditions more conducive and less conducive to the generation of hydrogen are known (Masukawa, Appl Microbiol Biotechnol 2002 Apr;58(5):618-24; Benneman JR . Proceedings of the 10th World Hydrogen Energy Conference, Cocoa Beach, FL, USA (1994) ; Papen, Biochimie 1986 Jan;68(1):121-32). Methods for culturing other bacteria such as E. coli under conditions more conducive and less conducive to the generation of hydrogen are known (Nandi, J Bacteriol 1985 Apr;162(1):353-60). The culture media may be solid or liquid.

[050] Standard growth media for other types of cells such as bacteria, cyanobacteria, and photosynthetic bacteria are known (see Maniatis et al. (1989) Molecular Cloning : A Laboratory Manual Cold Spring Harbor Laboratory;

Masukawa, Appl Microbiol Biotechnol 2002 Apr;58(5):618-24; and Papen et al., Biochimie 1986 Jan;68(1):121-32; Dzelzkalns, J Bacteriol 1986 Mar;165(3):964-71). Preferably the cells are cultured in liquid media during a screening or selection process since a desired strain that is capable of generating large amounts of hydrogen in the presence of oxygen is commercially deployed in liquid media.

Culturing Green Algae under conditions less conducive to the generation of hydrogen

[051] Green algae such as Chlamydomonas reinhardtii are grown in atmospheric conditions (ie: normal air), with or without illumination, according to standard protocols (Harris, (1989) The Chlamydomonas Sourcebook. Academic Press, New York; Rochaix J-D et al. (1998) The Molecular Biology of Chloroplasts and Mitochondria in Chlamydomonas (Advances in Photosynthesis, Vol 7). A culture is grown for any period of time under these conditions. Although it is desired to grow the cells overnight to obtain a healthy culture, if the starting cells were also grown under any conditions less conducive to the generation of hydrogen the culture need not be grown for a long periods of time. All that is necessary is for the cells to be cultured for some amount of time, preferably at least 5 minutes under conditions less conducive to the generation of hydrogen, before harvesting. More preferably, the cells are cultured for one or more hours before harvesting. Alternatively, cells are grown and then frozen. The exact conditions and duration of culturing are not vitally important, and trivial differences can be incorporated into the protocol, as long as the cells were not placed in conditions more conducive to the generation of hydrogen within at least about 10 minutes before harvesting. For example, the cells are cultured in Sager's minimal media or TAP media in light.

Culturing green algae under conditions more conducive to the generation of hydrogen

[052] In one example, green algae such as C. reinhardtii are cultured under conditions in which no sulfur is present in the media and atmospheric oxygen is not present in any gas space contacting the media. After about 15 hours under such conditions, green algae cells begin producing hydrogen. (Zhang, Planta (2002) Feb;214(4):552-61; Melis, Plant Physiol (2000) Jan;122(1):127-36). In other methods, cells are provided minimal amounts of sulfur, such as between 10 and 50 micromolar sulfur, and under such conditions cells generate hydrogen (Kosourov, Biotechnol Bioeng 2002 Jun 30;78(7):731-40).

[053] Preferably the cells are cultured in liquid media during a screening or selection process since a desired strain that is capable of generating large amounts of hydrogen in the presence of oxygen is commercially deployed in liquid media. In other words, it is desirable to screen or select for cells in the same type of media as will be used for commercial hydrogen production. For this reason liquid growth media is preferred. Growth media for Chlamydomonas cells, such as Sager's Minimal Media and Hunters Trace Element Media, are described in sources such as Harris E., (1989) The Chlamydomonas Sourcebook. Academic Press, New York and Rochaix J-D et al. (1998) The Molecular Biology of Chloroplasts and Mitochondria in Chlamydomonas (Advances in Photosynthesis, Vol 7). These growth media can be made as solid agar or as liquid. Other green algae media can be used, such as Tris-Acetate-Phosphate (TAP) media or Sueoka's media, as described in Harris and other sources. Minimal media such as Sager's (also known as Sager-Granick) is preferred when the host organism is or can be photoautotrophic because it is desirable to evolve microbes to generate hydrogen using only sunlight as energy. Sager's media is an example of photoautotrophic growth requiring media.

[054] Any component of the culture media may be manipulated. For example, a selection molecule such as an antibiotic is added to the culture media and a corresponding selectable marker gene is incorporated into the

transformation vector containing the recoded and recombined hydrogenase library.

[055] Optionally, other components of the culture media are manipulated such as amount of sulfur in the media. The level of sulfur may be increased, decreased, or held constant throughout the period of culture. (see Melis et. al. Plant Physiol (2000) Jan;122(1):127-36 and Zhang et al. Planta (2002) Feb;214(4):552-61).

[056] Another component that may be optionally added to the culture media is metronidazole (MNZ). MNZ is a strong oxidizer of reduced ferredoxin. Ferredoxin accepts electrons from the Photosystem I complex and transfers them to the hydrogenase to supply electrons for the $2H^+ + 2^{e-} \rightarrow H_2$ reaction. When MNZ is added to the culture media a controlled amount of oxygen is also added to the culture container and cells that survive are assayed for hydrogen production. In a typical experiment, C. reinhardtii cells that survive the MNZ treatment protocol; cultured for example in Saeger's minimal media in 20 mM MNZ; 1mM Sodium Azide; 2% oxygen; 200 W/m^2 light for 20 minutes, with expression of one or more mutagenized nucleic acid sequences, are placed in liquid culture media in multiwell plates and assayed for hydrogen production. It is unnecessary to count the number of independent transformants that survive the MNZ treatment. Any transformant that survives the treatment is capable of producing more hydrogen under a certain level of oxygen than a wild-type cell, and therefore all survivors are assayed for hydrogen production without regard to the number or percent of mutant survivors. For an example of the use of MNZ, see U.S. Patent 5,871,952.

[057] In one embodiment, cells are cultured in a Tris-acetate-phosphate media, at approximately pH 7.0 (Harris, (1989) The Chlamydomonas Sourcebook. Academic Press, New York). The cultures are bubbled with 3% CO_2 in air at 25°C. The cultures are continuously illuminated. After at least five minutes of culturing under these conditions, cells are harvested and are resuspended in the same media as before except for the absence of sulfur. The cells are then cultured under continuous illumination. Alternatively, the cells are originally cultured in the absence of acetate, but under continuous illumination (ie: photoautotrophically), and are then transferred to media that contains an absence of sulfur. Alternatively, culture conditions comprise culturing the cells in media that is devoid of sulfur, iron, or manganese, or any combination of these three elements.

[058] In another embodiment, frozen aliquots of green algae are thawed in culture media devoid of sulfur and continuously cultured, in the presence of light, for at least five minutes. The cells are then harvested.

[059] There are other culture conditions for some algae species that are conducive to the generation of hydrogen besides the sulfur deprivation method. For instance, blue-green algae produce hydrogen when starved of nitrogen (Weissman, Appl Environ Microbiol 1977 Jan;33(1):123-31). Hydrogen is also generated when green algae are cultured in the absence of light when the culture is flushed with gases, such as argon, that remove oxygen from the media (Happe, Eur J Biochem (2002) Feb;269(3):1022-32).

Generation of a differential expression profile: comparison of RNA between cells cultured in conditions more conducive to the generation of hydrogen and cells cultured in conditions less conducive to the generation of hydrogen

[060] Once at least two sets of cells are cultured under conditions more conducive and less conducive to the generation of hydrogen, RNA samples are extracted from the cells. Methods and protocols for the isolation of RNA from bacterial and algae cells are well known in the art (Maniatis et al. (1989) Molecular Cloning : A Laboratory Manual Cold Spring Harbor Laboratory; Harris, (1989) The Chlamydomonas Sourcebook. Academic Press, New York; Rochaix J-D et al. (1998) The Molecular Biology of Chloroplasts and Mitochondria in Chlamydomonas (Advances in Photosynthesis, Vol 7).

[061] The RNA is isolated from both the cells placed under conditions more conducive to the generation of hydrogen as well as cells placed under conditions less conducive to the generation of hydrogen. There is no requirement that both sets of cells be grown simultaneously or that RNA be isolated from both sets of cells simultaneously. There is also no requirement that the same strain of microbe be used in both culture conditions, although it is preferred that they be the same strain.

[062] After RNA is isolated from the cells, a plurality of methods can be utilized to generate a differential expression profile.

[063] In one embodiment, the RNA is placed on microarrays such as silicon chips or glass slides containing sequences corresponding to known sequences from the genome of the cells. It is not necessary that the sequences immobilized onto the microarray are derived from the same strain or species of the cells from which RNA are isolated as long as the genome of the cells used to make the microarray is somewhat homologous to the genome of the cells from which the RNA is isolated. For instance, the cells exposed to conditions more conducive and less conducive to the generation of hydrogen are Chlamydomonas fusca while the sequences immobilized on the microarrays are Chlamydomonas reinhardtii. Utilizing evolutionarily related strains of microbes for purposes of RNA isolation and microarray sequence immobilization provides reliable data, and the methods disclosed herein are utilized with a variety of microbes. RNA molecules isolated from cells hybridize with nucleic acid molecules immobilized on the microarray to form double stranded RNA duplexes. Such duplexes are detected by a variety of methods known in the art (such as the GeneChip® product and associated scanning techniques produced by Affymetrix Inc., Santa Clara, CA.; Dudley, Proc Natl Acad Sci U S A 2002 May 28;99(11):7554-9). In one embodiment the RNA isolated from cells is amplified by PCR and labeled nucleotides are incorporated into the newly synthesized nucleic acid molecules. These molecules are digested with a nuclease, denatured to single stranded molecules, and hybridized to the immobilized sequences on the chip. Double stranded duplexes that form contain the labeled nucleotides from the PCR reaction in one strand, and these duplexes are visualized. For example, the label incorporated into the molecules in the PCR reaction is a fluorescent molecule, and the microarray is placed into a fluorescence detection chamber. Such microarray technology is well known in the art. For instance, microarrays containing over 2,700 unique genes from C. reinhardtii are commercially available (Chlamydomonas Genome Project, Duke University, Durham, N.C.). In addition to the ability to visualize whether or not a duplex has formed on a particular spot corresponding to a particular gene on the chip, this technology also quantitates the difference in the amount of duplex formed on a given spot between two or more experiments using different RNA samples. This differentiation ability allows the identification of differentially regulated genes between cells grown in culture conditions more conducive to the generation of hydrogen and less conducive to the generation of hydrogen.

[064] Upon hybridization of the RNA samples from two or more sets of cells, genes that are upregulated or downregulated between the two sets of cells are identified. For example, the iron hydrogenase gene in Chlamydomonas is turned on when the cells are exposed to conditions more conducive to the generation of hydrogen, however the gene is turned off when the cells are exposed to conditions not conducive to the generation of hydrogen. When the two RNA samples are placed on microarrays containing immobilized sequences corresponding to the genome of C. reinhardtii, a spot on the chip containing the sequence of the iron hydrogenase gene contains a duplex of nucleic acid when the RNA sample is isolated from cells exposed to conditions more conducive to the generation of hydrogen, whereas the spot does not contain a duplex when the RNA sample is isolated from the cells exposed to conditions not conducive to the generation of hydrogen. The C. reinhardtii iron

hydrogenase gene is differentially regulated between cells exposed or not exposed to conditions more conducive to the generation of hydrogen, and therefore the gene is identified as differentially regulated.

Generation of a differential expression profile: Suppression Subtractive Hybridization between cells cultured in conditions more conducive to the generation of hydrogen and cells cultured in conditions less conducive to the generation of hydrogen

[065] In another embodiment, RNA is isolated from both sets of cells and is put through the Suppression Subtractive Hybridization PCR technique (Diatchenko, Proc Natl Acad Sci U S A 1996 Jun 11;93(12):6025-30; Happe, Eur J Biochem (2002) Feb;269(3):1022-32; commercially available kits are provided by Clontech Laboratories, Inc., Palo Alto, CA). In this technique transcripts from genes expressed in one sample (in this case the cells cultured under conditions more conducive to the generation of hydrogen) but not the other (in this case the cells cultured under conditions less or not conducive to the generation of hydrogen) are selectively amplified through the PCR method. Genes amplified through this technique are differentially regulated genes.

Generation of a differential expression profile: Two Dimensional gel electrophoresis between cells cultured in conditions more conducive to the generation of hydrogen and cells cultured in conditions less conducive to the generation of hydrogen

[066] A differential expression profile is created by subjecting protein samples from both sets of cells to two dimensional gel electrophoresis. This technique is well known in the art, and is optionally coupled with mass spectrometry techniques to aid in the identification of proteins (Arthur, Kidney Int 2002 Oct;62(4):1314-21). Spots indicating proteins on a gel from cells exposed to conditions more conducive to the generation of hydrogen but not present or present in different amounts on a gel from cells exposed to conditions less conducive to the generation of hydrogen correspond to proteins encoded by differentially regulated genes. Two dimensional gel electrophoresis analysis is advantageous for purposes such as monitoring the content of organelles such as chloroplast or multiprotein complexes such asphotosystem I that are involved in the production of hydrogen. (Dreger, Eur J Biochem. 2003 Feb;270(4):589-99).

Generation of a differential expression profile: Other Methods:

[067] In another embodiment, a differential expression profile is created by analyzing only a single gene or a small set of genes through methods such as Northern blotting, Western blotting, or activity assays specific to a protein of interest (Maniatis et al. (1989) Molecular Cloning : A Laboratory Manual Cold Spring Harbor Laboratory). A plurality of methods, specific to each gene, is employed to assess a difference in the activity of a gene or protein between two or more samples of cells exposed to different conditions. Any difference in conditions that a cell is exposed to may cause differential activity of some genes and/or proteins, including but not limited to components of culture media, temperature, exposure to sunlight or light of varying wavelengths, the presence of specific nutrients or elements, exposure to certain molecules, and exposure to other organisms or viruses.

Identification of differentially regulated genes

[068] After generation of the differential expression profile, any gene or protein demonstrated to be differentially regulated when cells are exposed to conditions more conducive to the generation of hydrogen versus conditions less conducive to the generation of hydrogen is a target for engineering efforts. For instance, the iron hydrogenase gene

in C. reinhardtii is differentially regulated between conditions more conducive to the generation of hydrogen and conditions less conducive to the generation of hydrogen.

[069] Also provided are methods for the identification of genes and proteins downregulated when cells are exposed to conditions more conducive to the generation of hydrogen. Such genes are targets for mutation, deletion from the genome, or downregulation through methods such as RNA interference. Alternatively, molecules capable of inhibiting the activity of proteins downregulated when cells are exposed to conditions more conducive to the generation of hydrogen are added to the culture in order to stimulate the cells to generate an increased amount of hydrogen.

Providing mutagenized nucleic acid sequences corresponding to differentially regulated genes

[070] Clones of genes identified as differentially regulated are obtained. Creation of full-length cDNA molecules is standard in the art (Maniatis et al. (1989) Molecular Cloning : A Laboratory Manual Cold Spring Harbor Laboratory), however gene fragments are also used. The gene or gene fragment is mutagenized using one or more mutagenesis methods.

[071] In one embodiment, the gene is amplified using error-prone PCR. Error-prone PCR is a standard procedure in the art (Leung, Technique (1989) 1, 11-15). In this technique the gene of interest is amplified using a DNA polymerase under conditions that are deficient in the fidelity of replication of sequence. The result is that the amplification products contain at least one error in the sequence. When a gene is amplified and the resulting product(s) of the reaction contain one or more alterations in sequence when compared to the template molecule, the resulting products are mutagenized as compared to the template.

[072] Alternatively, the gene of interest is cloned into a suitable vector and used to transform a microbe. The microbe is then grown while exposed to a mutagenizing agent such as nitrosoguanidine or ethyl methanesulfonate (Nestmann, Mutat Res 1975 Jun;28(3):323-30), and the vector containing the gene is then isolated from the host.

[073] In one embodiment, the gene identified as upregulated is mutagenized through gene reassembly, saturation mutagenesis, or other directed evolution techniques. These techniques are known in the art (U.S. Patent 5,605,793, U.S. Patent 5,830,721, U.S. Patent 6,165,793, U.S. Patent 6,180,406, U.S. Patent 5,939,250, U.S. Patent 6,171,820, U.S. Patent 6,361,974, U.S. Patent 6,358,709, U.S. Patent 6,352,842, U.S. Patent 6,238,884, U.S. Patent 6,420,175, U.S. patent 6,287,861 and related patents; Coco et al., Nat Biotechnol 2001 Apr;19(4):354-9).

[074] It is preferable but not necessary that nucleic acid molecules used in shuffling protocols use the same codon to encode each individual amino acid. For example, even though 6 different amino acids encode Arginine, only CGC is used. It is also preferable that the codon used to encode each amino acid is the most preferred codon in an organism that is transformed with the shuffled sequences. Using only one codon that is the most preferred codon in the organism is preferred because it allows the nucleic acid fragments to anneal better because they have higher nucleotide sequence identity. In addition, every protein encoded by a shuffled sequence is translated at equal efficiency by the organism. In one embodiment, the organism is C. reinhardtii, at least nucleic acid molecule encoding one segment of a protein from SEQ ID NOs: 1-112 is used in a shuffling protocol, and the nucleic acid molecules that are used in the shuffling protocol use only the most preferred codon from C. reinhardtii as depicted in figure 10.

[075] In one embodiment, the differentially regulated gene is digested with a nuclease such as Dnase I to form random fragments. These fragments are mixed with similarly digested fragments of at least one other gene that contains some sequence homology to the differentially regulated gene. Alternatively the fragments are pooled with

synthetic single or double stranded oligonucleotides corresponding to sequences from genes possessing homology or partial homology to the differentially regulated gene. The mixed fragments are denatured to form single stranded molecules and the molecules are then allowed to anneal to each other. The fragments are put through an extension protocol such as primerless PCR in which 3' ends of fragments are extended through the use of a DNA polymerase enzyme. The resulting mixture contains a library of shuffled sequences that are used to transform cells for screening or selection procedures.

[076] In one embodiment genes that are homologous to genes that are (a) identified as differentially regulated and (b) are further identified as upregulated when cells are exposed to conditions more conducive to the generation of hydrogen are isolated from evolutionarily similar microbes. For example, the iron hydrogenase gene is upregulated in C. reinhardtii when the cells are exposed to conditions more conducive to the generation of hydrogen. Other iron hydrogenase genes are isolated from microbes that are evolutionarily related and/or are known to possess an iron hydrogenase gene. For sequences of genes homologous to the gene identified as differentially regulated that are already known, gene fragments corresponding to these genes may be chemically synthesized using known sequence information; it is not necessary that such genes be actually cloned from their natural source in order to be utilized in shuffling experiments. Examples of such known iron hydrogenase genes include those listed in the sequence listing.

[077] In one embodiment, nucleic acid fragment encoding proteins sequences of at least 5 amino acids are used in shuffling experiments. Alternatively, the fragments encode at least 6 amino acids, and in some instances at least 8 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 or more amino acids.

[078] These genes are isolated through procedures known in the art. For instance, the C. reinhardtii iron hydrogenase gene is used as a probe to screen cDNA or genomic DNA libraries of other green algae. In particular, the highly conserved "H-cluster" sequence corresponding to the active site of iron hydrogenases is used as a probe (Peters, Science (1998) Dec 4;282(5395):1853-8, Nicolet, Structure Fold Des (1999) Jan 15;7(1):13-23). Alternatively, PCR primers corresponding to sequences from the C. reinhardtii iron hydrogenase gene are used to amplify iron hydrogenase genes from other microbial genomes. In this method the PCR template is genomic DNA, a cDNA library, or RNA for use in RT-PCR. The sequences isolated from each microbe are mixed and put through a shuffling procedure.

[079] In one embodiment, a plurality of genes is identified from the differential expression profile as upregulated when C. reinhardtii cells are exposed to conditions more conducive to the generation of hydrogen. Sequence information from these genes is used to generate probes and PCR primers corresponding to the sequences. A plurality of green algae species, originally isolated from disparate geographic locations, are cultured under conditions more conducive to the generation of hydrogen. A cDNA library from each green algae species is generated and utilized for the isolation of sequences corresponding to each of the sequences identified from C. reinhardtii as differentially regulated using the probes corresponding to the upregulated C. reinhardtii sequences. The isolated gene sequences are used for shuffling.

[080] In one embodiment, the plurality of genes is shuffled in reactions containing synthetic chimeric oligonucleotides. The chimeric oligonucleotides possess on one end sequence corresponding to either the 5' or 3' end of the coding region of genes included in the shuffling reaction. On the other end these chimeric oligonucleotides contain heterologous sequence, such as unique sequences not found in the genes that are shuffled or in the genome of the hydrogen producing microbe. The unique sequences are used to connect different components of DNA constructs containing mutagenized nucleic acid sequences (Figure 3). Other chimeric oligonucleotides contain sequences corresponding to (a) a promoter sequence and (b) a unique sequence. The sense

and antisense strands of unique sequences are used to join mutagenized nucleic acid sequences with promoter sequences and other types of sequence heterologous to the mutagenized nucleic acid sequences. For example, a promoter sequence imparts transcriptional activation to a downstream mutagenized nucleic acid sequence when placed in a Chlamydomonas cell that is exposed to light (Hahn, Curr Genet (1999) Jan;34(6):459-66; Loppes, Plant Mol Biol 2001 Jan;45(2):215-27; Villand, Biochem J 1997 Oct 1;327 (Pt 1):51-7). Other light-inducible promoter systems may also be used, such as the phytochrome/PIF3 system (Shimizu-Sato, Nat Biotechnol 2002 Oct;20(10):1041-4). Alternatively or in addition, the promoter sequence imparts transcriptional activation to a downstream gene when placed in a Chlamydomonas cell that is exposed to light and heat (Muller, Gene (1992) Feb 15;111(2):165-73; von Gromoff, Mol Cell Biol (1989) Sep;9(9):3911-8). Alternatively the promoter sequence imparts transcriptional activation to a downstream gene when an exogenous molecule is added to the culture media using receptors not present in the wild-type cell such as receptors for estrogen, ecdysone, or others (Metzger, Nature 1988 Jul 7;334(6177):31-6; No, Proc Natl Acad Sci U S A 1996 Apr 16;93(8):3346-51). Alternatively the promoter sequence imparts transcriptional activation in a constitutive fashion, such as the promoter of the psaD gene (Fischer, WO 01/48185). When the shuffled gene fragments are annealed and subjected to primerless PCR, the 5' and 3' ends of the shuffled coding regions anneal to chimeric oligonucleotides that in turn anneal to other heterologous sequences such as promoters and 3' untranslated regions that enhance expression levels (Lumbreras, Plant J (1998) 14(4): 441-447). The 5' end of every coding sequence created through the shuffling procedure is annealed to a chimeric oligonucleotide corresponding to a unique sequence. The unique sequence in turn anneals to a nonshuffled segment of DNA containing a promoter sequence (Figures 3, 4). Unique sequences are thus used to attach components of DNA constructs to each other that do not possess sequence homology. In addition, chimeric oligonucleotides are included that possess homology to internal parts of the coding region of shuffled genes as well as intron sequences to direct the insertion of intron sequences into coding regions to aid in effective expression levels (Lumbreras, Plant J (1998) 14(4): 441-447).

[081] Chimeric oligonucleotides may be used to connect any part of a nucleic acid construct to another in shuffling protocols. Intron, transcriptional terminator, splice sequences, centromeres, selectable and screenable markers are all introduced into nucleic acid constructs through annealing these elements to chimeric oligonucleotides that contain heterologous sequence, followed by promoterless PCR protocols.

[082] In one embodiment, libraries of individually shuffled homologous genes with unique sequences at each end are mixed with other distinct libraries of individually shuffled homologous genes that also contain unique sequences at both 5' and 3' ends. Also mixed with the shuffled libraries of coding sequences are nonshuffled segments containing structural and functional DNA elements such as promoters, 3' untranslated regions, and screenable or selectable markers. The nonshuffled segments of DNA are also flanked with unique sequences, all of which are identical to unique sequences flanking certain shuffled sequences. All of the molecules are denatured, annealed, and subjected to a primerless PCR reaction in which "sense" and "antisense" unique sequences anneal to each other and prime extension by a polymerase, thus placing each shuffled and nonshuffled sequence into its desired place on the resulting DNA construct. The resulting library of DNA constructs contains shuffled genes operatively linked to promoter sequences. (Figures 3, 4)

[083] In one embodiment chimeric oligonucleotides contain sequence corresponding to genes being shuffled and heterologous sequence corresponding to introns, splice sequences, centromeres, selectable markers, unique sequences or other linker sequences designed to serve as structural parts of the construct. The design of the DNA construct using these chimeric oligonucleotides creates a functional DNA construct directly from the shuffling

procedure. Any desired component of a DNA construct is included through the use of chimeric oligonucleotides that connect heterologous sequences of the construct during the annealing step. For instance, the inclusion of a light-inducible promoter allows the shuffled versions of differentially regulated genes to be activated by light rather than the conditions more conducive to the generation of hydrogen.

[084] In one embodiment each DNA construct in the library of DNA constructs contains a plurality of shuffled genes that possess sequence homology to a set of upregulated differentially regulated genes. Each coding region has an upstream light-inducible promoter and a downstream untranslated transcriptional terminator sequence. Each coding region contains an intron and functional splice sequences. Each construct contains at least one selectable marker. Constructs optionally also contain other functional or structural sequences. For example, centromeres or other sequences employed for the purpose of allowing the construct to be retained in dividing cells and/or sequences that aid in integration of the construct into random or specific regions of the host genome are included in the construct. In other embodiments the promoter is constitutive or is inducible by a stimulus other than light, such as the addition of a small molecule to the culture media.

[085] In one embodiment, DNA constructs are used to turn off or downregulate the expression of differentially regulated genes that are downregulated when cells are exposed to conditions more conducive to the generation of hydrogen. These constructs work through the use of antisense and/or RNA interference methods. In this embodiment, a DNA construct containing at least one antisense sequence operatively linked to a promoter is used to transform cells for the purpose of downregulating the expression of a gene or genes that are naturally downregulated when cells are exposed to conditions more conducive to the generation of hydrogen. For example, in Chlamydomonas, antisense inhibition is utilized to effect a drop in expression of the targeted gene (Schroda, Plant Cell (1999) Jun;11(6):1165-78). Alternatively, an RNA interference (RNAi) construct is used (Fire, Nature (1998) Feb 19;391(6669):806-11 ; Fuhrmann, J Cell Sci (2001) Nov;114(Pt 21):3857-63). In one embodiment, DNA constructs are synthesized that contain shuffled sequences corresponding to genes upregulated when cells are exposed to conditions more conducive to the generation of hydrogen and RNAi sequences corresponding to genes downregulated when cells are exposed to conditions conducive to the generation of hydrogen. Both the shuffled sequences and the RNAi sequences are functionally coupled to promoters that are activated by the same stimuli, different stimuli, or are constitutively active.

[086] In one embodiment genes downregulated when cells are exposed to conditions less conducive to the generation of hydrogen are removed from the genome through gene targeting methods that utilize homologous recombination (Naver, Plant Cell 2001 Dec;13(12):2731-45).

[087] In one embodiment molecules that interfere with the function of proteins that are encoded by genes downregulated when cells are exposed to conditions more conducive to the generation of hydrogen are either placed in the culture media or synthesized by proteins encoded by transgenes inserted into cells.

[088] In one embodiment the DNA constructs containing shuffled upregulated differentially regulated genes contain genes encoding screenable or selectable markers at each end of a linear DNA construct. For example, at one end of the construct is a gene encoding a fluorescent protein optimized for use in Chlamydomonas (Fuhrmann, Plant J (1999) Aug;19(3):353-61). At the other end is a gene encoding a selectable marker gene that imparts resistance to an antibiotic (Stevens, Mol Gen Genet (1996) Apr 24;251(1):23-30). Between the fluorescent protein and the antibiotic resistance gene are shuffled versions of genes upregulated when cells are exposed to conditions more conducive to the generation of hydrogen or are involved in the hydrogen production pathway, such as ferredoxin, catalase, isoamylase, malate dehydrogenase, 14-3-3 protein, enolase, aldolase, ribosomal protein S8,

ribosomal protein L17, ribosomal protein S18, ribosomal protein L37, ribosomal protein L12, ribosomal protein S15, iron-hydrogenase, and components of the photosystem I, photosystem II and cytochrome b_6-f complexes. Components of the photosystem I and II complexes are disclosed, for example, in Elrad, Curr Genet. 2003 Dec 2. Hydrogen can be produced in C. reinhardtii for example, by pathways that opetare in light and dark. Mutagenized genes from either pathway can be assayed using the methods disclosed herein. Cells are transformed with the library of constructs and are cultured in media containing the antibiotic. Cells that survive under these culture conditions are run through a fluorescence activated cell sorter that plates each cell expressing the green fluorescent protein onto a grid pattern on solid media or into multiwell plates containing liquid growth media containing the antibiotic. Colonies are screened or selected for the ability to generate an increased amount of hydrogen. Cells that retain both markers have also retained all the sequence in the DNA construct between the two markers. Large numbers of genes may be placed between the two markers. Preferably only cells that retain both markers are put through screening or selection procedures.

[089] In one embodiment the mutagenized nucleic acid sequence encodes an iron hydrogenase protein and the cell is a green algae species such as C. reinhardtii. Further, the mutagenized nucleic acid sequence is generated by mutagenizing a C. reinhardtii iron hydrogenase gene at at least one amino acid position. The mutagenized nucleic acid sequence is used in a construct to transform the cell. Preferably, the iron hydrogenase protein retains the capacity to functionally interact with a ferredoxin or other electron donor in the cell. "Functionally interact" means that a ferredoxin or other electron donor transfers electrons to the hydrogenase protein. Preferably the sequence change(s) caused by the mutagenesis of the C. reinhardtii iron hydrogenase gene does not disrupt the functional interaction between the protein encoded by the mutagenized C. reinhardtii iron hydrogenase gene and ferredoxin or another electron donor. Preferably the mutagenesis creates an oxygen tolerance phenotype without disrupting the functional interaction with a ferredoxin. More preferably, the mutagenesis creates an oxygen tolerance phenotype while enhancing the functional interaction with a ferredoxin. An example of an enhanced functional interaction with ferredoxin is a functional interaction that allows more electrons to be shuttled from the endogenous ferredoxin to the mutagenized iron hydrogenase per unit time under than with the non-mutagenized C. reinhardtii iron hydrogenase. An enhanced functional interaction can also be screened or selected for by mutagenizing the ferredoxin, as described in Example 2.

Providing mutagenized nucleic acid sequences corresponding to genes known to be involved in a hydrogen production pathway

[090] Wild type iron hydrogenase genes are preferred mutagenesis targets with which to generate mutagenized nucleic acid sequences. Mutagenesis preferably alters characteristics such as oxygen tolerance while not altering characteristics such as the ability to functionally interact with ferredoxin.

[091] In one embodiment, the C. reinhardtii iron hydrogenase gene is mutated to alter amino acid residues in and near the gas channel. The gas channel is a section of iron hydrogenases, depicted in figure 9, that allows newly formed hydrogen molecules to leave the protein. Oxygen irreversibly inactivates the active site of iron hydrogenases by entering the active site through the gas channel (for background see Ghirardi, Appl Biochem Biotechnol (1997) 63-65: 141-151). Because hydrogen molecules are smaller than oxygen molecules, narrowing the gas channel using methods deiclosed herein provides iron hydrogenases that are not inactivated by oxygen. Preferably, substitutions of residues that are in and near the gas channel generate side chains that are of higher molecular weight or are longer than the side chain at that position in the wild type protein. Such substitutions are

preferable because they narrow the gas channel and block the entry of oxygen into the active site. As one nonlimiting example, residues in the highly conserved $X^1X^2X^3FX^4X^5X^6GGVMEAAX^7R$ segment can be mutated. This segment forms a turn followed by an alpha helix. The F corresponds to Phe234 in the wild type C. reinhardtii iron hydrogenase. The X residues are highly variable between iron hydrogenase from different species. For example, the $X^4X^5X^6$ residues are GVT, GAT, GVS, GNS, CAS, and numerous other sequences in different iron hydrogenases. Nonetheless, members of the iron hydrogenase family usually have a G as the first residue of this triplet. Although the GGVMEAA amino acid motif is highly conserved among members of the iron hydrogenase family, there are some iron hydrogenases that have variant sequences corresponding to this motif. For example, the D. fructosovorans iron hydrogenase (GenBank Accession number D57150) has the sequence GGVIEAA. Thus, even highly conserved motifs that surround the gas channel are tolerant of change.

[092] Other amino acid motifs also form secondary structures near the gas channel. For example, the ADX^8TIX^9EE motif is in close contact with the channel. In particular, the T, I and X^9 residues are near the channel.

[093] In one embodiment, highly variable amino acids are subjected to saturation mutagenesis. In another embodiment, highly variable amino acids are substituted with any amino acid that is of a higher molecular weight hat the wild type amino acid at that position in either of the C. reinhardtii iron hydrogenases. In another embodiment, variable amino acids in either of the C. reinhardtii iron hydrogenases are substituted with amino acids that are found in the corresponding position in iron hydrogenases from different species. In yet another embodiment, the $X^1X^2X^3FX^4X^5X^6GGVMEAAX^7R$ motif is mutated in either of the C. reinhardtii iron hydrogenases referred to as hydA and hydB (Forestier, Eur J Biochem. 2003 Jul;270(13):2750-8), wherein some of the X residues are substituted with amino acids that are found in the corresponding position in iron hydrogenases from different species while other X residues are substituted with residues that are no found in any known species. In one embodiment residues $X^1X^2X^3$ are from species 1, residues $X^4X^5X^6$ are from species 2, and residue X^7 is from species 3, where these X residues are placed in the context of a C. reinhardtii iron hydrogenase protein, and where none of species 1, 2, or 3 is C. reinhardtii. The methods provided herein include mutagenizing genes by substituting any segment of a protein sequence into another protein sequence, including genes encoding iron and nickel-iron hydrogenase proteins. Preferable lengths for segments include 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 or more amino acids. Of course, the methods provided also included substituting single amino acids from one species into the proteins of another species at a particular position as well as substituting amino acids that do not correspond to amino acids of another species at a particular position.

[094] In another embodiment, gene reassembly of the iron hydrogenase is performed. Sections of the C. reinhardtii iron hydrogenase active site region that are both highly conserved and correspond to the gas channel are used to construct a library of iron hydrogenase genes, depicted schematically in figure 13. In step 1, the library of iron hydrogenase amino acid sequences from SEQ ID NOs: 1-112 was aligned using sequence manipulation software (DS Gene, Accelyrys Inc., San Diego, CA). The key in figure 15 shows the identity of amino acids from step 1 and codons from steps 2-9. All bars in steps 2-9 correspond to codons that encode the amino acids from the bars of step 1. Each bar in steps 2-9 therefore depicts a codon triplet of oligonucleotide sequence. In step 2, conserved amino acid segments were identified in the alignment and reverse-translated into single stranded oligonucleotide sequences utilizing C. reinhardtii most preferred codons. In step 3, 3 codons encoding amino acids flanking these highly conserved gas channel sequences were re-written as the C. reinhardtii flanking sequence of the oligonucleotides. Even though these oligonucleotides encode different gas channel segments from the C. reinhardtii iron hydrogenase, the combination of the recoding process and the substitution of3 flanking C. reinhardtii codons

generates enough nucleotide similarity that these oligonucleotides anneal to a complementary strand encoding the recoded, wild-type C. reinhardtii iron hydrogenase. In step 4, the set of recoded oligonucleotides corresponding to diverse gas channel segments are annealed to a single stranded DNA molecule that encode C. reinhardtii Iron hydrogenase protein using the same C. reinhardtii most preferred codons. In addition, oligonucleotides corresponding to wild type C. reinhardtii amino acid sequences with single residue substitutions designed to narrow the gas channel can also be included in the annealing reaction. A C. reinhardtii C-terminal primer is also added to the annealing reaction. The single stranded molecule is generated by isolating the gene from a plasmid grown in a methylating host cell, followed by denaturation and separation of the strands by HPLC or other standard procedures, as described for example in U.S. patent 6,361,974. As shown in step 5 of figure 14, different combinations of segments anneal to each full length complementary strand. Addition of DNA Polymerase in step 6 extends the annealed oligonucleotides, creating a library of double stranded hybrid molecules with mismatches at "context" residue positions. Preferably the DNA Polymerase is exonuclease-deficient to prevent it from degrading parts of annealed primers in its path as it extends between annealed primers. In step 7, the methylated strands are digested using a methylation-sensitive endonuclease, as described for example in U.S. patent 6,361,974. In steps 8-9, N-terminal C. reinhardtii primer and DNA Polymerase are added to the library of novel iron hydrogenase molecules. As an alternative to methylation, the C-terminal primer shown first in step 4 can be biotinylated, and the mismatched wild type and library strands can be separated in step 7 by denaturation and separation using immobilized streptavidin.

[095] The result of the above process is a library of double stranded iron hydrogenase sequences that have random combinations of functional gas channel segments and C. reinhardtii framework/hinge regions. The population is cloned into C. reinhardtii cells and assayed as described in previous sections. This method does not use an exonuclease such as mung bean nuclease. No single stranded fragments that anneal to the methylated strand have partially overlapping binding sites. The advantage of this method of creating mutagenized nucleic acid sequences is that the library can be tested for oxygen tolerance but preserves C. reinhardtii framework/hinge domains that functionally interact with ferredoxin than a library made using other gene reassembly procedures such as the procedure shown in figures 2-3 that involves reassembly of the entire gene sequence. In a preferred embodiment, single stranded nucleotide molecules, using C. reinhardtii most preferred codons, encoding segments or fragments of segments depicted in figure 16 are used in the procedure. Although figure 17 depicts one possible arrangement of three diverse oligonucleotides that can be annealed to a single stranded wild type sequence, mixing oligonucleotides corresponding to each of the identified gas channel segments from SEQ ID Nos: 124-147 that have C. reinhardtii flanking codons produces a large number of possible combinations of library sequences. Each possible combination corresponds to a different gas channel architecture that can be tested for the ability to allow flow of hydrogen but not oxygen.

[096] Alternatively, other genes involved in a hydrogen production pathway are mutagenized. Examples of these genes are recited elsewhere in this application. As one example, genes encoding light antenna complexes are mutagenized and inserted into cells. For example, one or more genes from a light harvesting complex of C. reinhardtii, such as those disclosed in Teramoto, Plant Cell Physiol. 2001 Aug;42(8):849-56. (corresponding to GenBank accession numbers M24072, AF104630, AF104631, AB050007, X65119), and Elrad, Curr Genet. 2003 Dec 2 (lhcbm1, lhcbm2, lhcbm3, lhcbm4, lhcbm5, lhcbm6, lhcbm8, lhcbm9, lhcbm11, lhca1, lhca2,lhca3, lhca4, lhca5, lhca6, lhca7, lhca8, lhca9, lhcb4, lhcb5, lhcq, l1818-111818-2, elip1, elip2, elip3, elip4, and elip5) are mutagenized and used to transform C. reinhardtii. Transformants are screened or selected for the ability to produce

an increased amount of hydrogen under conditions such as high light, low light, sunlight, or light of a certain wavelength range. For example, segments of amino acids from antenna proteins of one species are inserted into antenna proteins from C. reinhardtii. The mutagenized nucleic acid sequence is then inserted into C. reinhardtii cels and the transformed cells are screened or selected for the ability to live and/or produce hydrogen in the presence of photoautotrophic media and light. In one embodiment the light is of a wavelength that wild type C. reinhardtii antenna proteins are not capable of harvesting.

[097] In another embodiment, an siRNA construct is used to transform a cell, where the siRNA construct is designed to reduce or eliminate the expression of a gene that reduces the photosynthetic efficiency or rate. For example, the C. reinhardtii lhcbm1 gene is reduced or eliminated in expression using siRNA (sequence of lhcbm1 in Elrad, Plant Cell. 2002 Aug;14(8):1801-16).

[098] In one embodiment, cell transformed with mutagenized antenna genes are cultured in the presence of light outside the normal wavelength range of the starting strain. For example, genes encoding purple bacteria antenna complexes are transformed into green algae such as C. reinhardtii. The genes include preferably only the most preferred codon of C. reinhardtii for each amino acid. Preferably, bacteriochlorophyll molecules are present in the cells, either synthesized by enzymes also present in the C. reinhardtii cell or added exogenously to the culture media. The cells are cultured in photoautotrophic media under light of wavelengths that wild type green algae are not capable of capturing, such as 770-920nm. Narrow ranges can be used as well, such as 800-900nm. In one embodiment, the α peptides of Rs. rubrum, Rb sphaeroides, and Rb. capsulatus are reverse translated into C. reinhardtii most preferred codons (see sequences from Davis, Biochemistry. 1997 Mar 25;36(12):3671-9.). These α peptide genes, encoding amino acids only in C. reinhardtii most preferred codons, are shuffled. The β peptides from the above three organisms, also as shown in Davis, are also reverse translated into C. reinhardtii most preferred codons and shuffled. The shuffled α and β peptides are cloned into expression vectors and used to transform C. reinhardtii. Preferably the α and β peptide sequences also include targeting domains that cause the expressed proteins to be embedded in light harvesting complexes of the C. reinhardtii thylakoid membrane. The transformed population is cultured under light of a wavelength above 700nm, preferably above 750 nm, more preferably above 800nm. Surviving strains are then assayed for hydrogen production in light of a wavelength above 700nm, preferably above 750 nm, more preferably above 800nm.

[099] In another embodiment, shuffling is performed using nucleic acid molecules encoding nickel-iron hydrogenase proteins, such as those in SEQ ID NOs: 113-122. Because these Ni-Fe hydrogenases are made of alpha and beta subunits, preferably the nucleic acid molecules encoding segments of each protein are shuffled in separate reactions. The shuffled libraries are expressed in cells that possess Ni-Iron hydrogenaserogenase maturation enzymes, such as E. coli.

Transforming cells with mutagenized nucleic acid sequences

[100] Cell transformation methods and selectable markers for photosynthetic bacteria and cyanobacteria are well known in the art (Wirth, Mol Gen Genet 1989 Mar;216(1):175-7; Koksharova, Appl Microbiol Biotechnol 2002 Feb;58(2):123-37; Thelwell). Transformation methods and selectable markers for use in bacteria are well known (Maniatis et al. (1989) Molecular Cloning : A Laboratory Manual Cold Spring Harbor Laboratory).

[101] In green algae, the nuclear, mitochondrial, and chloroplast genomes are transformed through a variety of known methods. (Kindle, J Cell Biol (1989) Dec;109(6 Pt 1):2589-601; Kindle, Proc Natl Acad Sci U S A (1990) Feb;87(3):1228-32; Kindle, Proc Natl Acad Sci U S A (1991) Mar 1;88(5):1721-5; Shimogawara, Genetics (1998)

Apr;148(4):1821-8; Boynton, Science (1988) Jun 10;240(4858):1534-8; Boynton, Methods Enzymol (1996) 264:279-96; Randolph-Anderson, Mol Gen Genet (1993) Jan;236(2-3):235-44).

[102] Selectable markers for use in Chlamydomonas are known, including but not limited to markers imparting spectinomycin resistance (Fargo, Mol Cell Biol (1999) Oct;19(10):6980-90), kanamycin and amikacin resistance (Bateman, Mol Gen Genet (2000) Apr;263(3):404-10), zeomycin and phleomycin resistance (Stevens, Mol Gen Genet (1996) Apr 24;251(1):23-30), and paromycin and neomycin resistance (Sizova, Gene (2001) Oct 17;277(1-2):221-9).

[103] Screenable markers are available in Chlamydomonas, such as the green fluorescent protein (Fuhrmann, Plant J (1999) Aug;19(3):353-61) and the Renilla luciferase gene (Minko, Mol Gen Genet (1999) Oct;262(3):421-5). Fluorescent proteins are also available for prokaryotic organisms.

[104] In one embodiment, libraries of gene sequences that encode proteins that physically interact are shuffled. Nucleic acid constructs are used for transformation procedures that contain a shuffled version of each gene. Sequences that encode proteins that interact in ways more conducive to the generation of hydrogen are screened or selected for. By mutagenizing sequences encoding proteins that physically interact, more favorable interactions are generated that lead to the production of increased levels of hydrogen. Examples of such proteins in the hydrogen production pathway that physically interact are iron-hydrogenase/ferredoxin and proteins in the photosystem I, photosystem II, and cytochrome b_6-f complexes. It is advantageous but not necessary to use pairs or sets of genes that encode proteins that physically interact from the same organisms. Providing interacting pairs or sets in the shuffling procedure increases the odds of obtaining favorable functional interactions due to the possibility of obtaining shuffled sequences on the same test construct that contain complementary interaction domains from the same organism, regardless of the sequence flanking either side of the interaction domain in any of the sequences.

[105] In one embodiment, a library of sequences corresponding to at least one mutagenized nucleic acid sequence derived from a differentially regulated gene is inserted into cells through a transformation procedure. Cells that have been transformed with the library are then put through a screening or selection process in which the cells are assayed for the ability to generate an increased amount of hydrogen when compared to the non-transformed strain or the strain transformed with only vector and/or screenable/selectable marker sequences.

Screening or Selecting for a Cell that Generates an Increased Amount of Hydrogen

[106] Cells are screened for the ability to produce hydrogen by a variety of methods. One method involves the use of gas chromatography, which is a well known method of detecting gases such as hydrogen. An intake device attached to the gas chromatography machine is placed in close enough proximity to the cell culture container or plate that it can detect, and preferably quantify, the hydrogen produced by the cells (U.S. Patent 5,100,781).

[107] Oxygen content may be manipulated in the culture container. The amount of oxygen in the culture container may be directly adjusted through gas exchange or indirectly by allowing or inducing the water-splitting mechanism of photosynthesis. The oxygen content, like all other culture parameters, may be manipulated throughout the culture period or held constant. The presence of some amount of oxygen is preferred if MNZ is added to the culture media. Preferred hydrogenase genes are capable of catalyzing the production of hydrogen in the presence of oxygen. A preferable amount of oxygen in a culture of commercially deployed cells for hydrogen production is an atmospheric level such as approximately 21%. Several rounds of screening or selection may be performed in which the oxygen content of the culture container may be increased between each successive round while hydrogen production is assayed. For example, a culture is exposed to 5% oxygen in the first screening or selection round,

10% oxygen in the second screening or selection round, 15% oxygen in the third screening or selection round, and 20% oxygen in the fourth screening or selection round. Other levels of oxygen that can be tested include more than 0.5%, more than 5.0%, more than 10%, more than 15%, approximately 21%, more than 21%, more than 25%, more than 30% or more than 35%.

[108] In one embodiment, the screening assay is a chemochromic film that turns from transparent to opaque in the presence of hydrogen. The assay is performed by placing films over arrays of multiwell plates containing libraries of C. reinhardtii transformants. As shown in figure 8, independent transformants are cultured in multiwell plates. The film seals each well. Hydrogen produced by cells is reversibly coordinated to the transition metal in the film, causing the film to go from transparent to opaque in a quantitative fashion. The film is photographed with digital imaging equipment and cells from wells corresponding to spots darker than the starting strain are selected for further rounds of mutagenesis.

[109] The assay is performed using a platform in which a variety of parameters are manipulated. The platform contains an enclosed chamber in which multiwell plates are exposed to a controlled gas environment. Lights are positioned over the chamber such that daylight/nighttime conditions may be mimicked. The temperature of the chamber may be manipulated corresponding to colder nighttime temperatures followed by warmer daytime temperatures. The platform allows the directed evolution procedure to create novel microbe strains that are best suited for commercial deployment. For example, in one embodiment strains that can produce hydrogen for hundreds of hours using constant light at a constant temperature are assayed for; in a second embodiment strains capable of producing large amounts of hydrogen during a warmer 12 hour light period after being exposed to a colder 12 hour dark period are assayed for. Strains produced by the second embodiment are best suited for commercial deployment because they are best able to conserve energy at night when the photosynthetic electron transport chain is not functional.

[110] In one embodiment, the hydrogen production assay mimics commercial deployment conditions through the use of deep-well plates made from non-transparent plastic material. When mutants are assayed for hydrogen production, the light available to the cells comes only from directly above the plates, mimicking conditions under which cells in a large bioreactor are exposed to light. Mutations that attenuate phototaxis (swimming towards light) under bright light conditions (but not dim conditions) prevent cells from accumulating at the surface of the media and blocking photons from penetrating deeper into the media. Mutations in the antenna complexes also enhance photon utilization efficiency.

[111] In one embodiment, cells transformed with mutagenized nucleic acid sequences are cultured under conditions in which gas in the culture container comprises 5% oxygen. Cells that generate an increased amount of hydrogen are recovered and mutagenized nucleic acid sequences are recovered from the cells. The mutagenized nucleic acid sequences are put through a further mutagenesis round and are used to transform cells. The transformed cells are cultured under 21% oxygen. Mutagenized nucleic acid sequences corresponding to differentially regulated genes whose wild type sequence encodes proteins that do not function or minimally function in atmospheric oxygen levels, such as the C. reinhardtii iron hydrogenase, provide oxygen tolerant variants to the transformed cells. Shuffling protocols that include versions of genes that possess desirable characteristics, such as the iron hydrogenase gene from Desulfovibrio vulgaris, which is reversibly inactivated by oxygen, are likely to generate shuffled genes with multiple desirable characteristics from different parent genes.

[112] In one embodiment cells transformed with mutagenized nucleic acid sequences are cultured in the presence of metronidazole and are selected for the ability to produce increased amounts of hydrogen according to known

methods (U.S. Patent 5,871,952).

[113] Alternatively other sensing methods are utilized. Compounds that reversibly react with hydrogen are used to synthesize films that are placed either directly on or in proximity to distinct colonies on culture plates or culture containers. The film changes a detectable characteristic in the presence of hydrogen, such as a change of color or a change from clear to opaque. In one embodiment, a substrate containing a hydrogen-dissociative catalyst metal such as tungsten trioxide is placed on or near colonies of cells and turns from transparent to blue/opaque in the presence of hydrogen (U.S. Patent 6,277,589).

[114] There are other methods, both direct and indirect, that are used to detect hydrogen, such as spectroscopic methods (U.S. Patent 6,309,604). Other types of gas sensors suitable for detection of hydrogen are well known in the art.

[115] Colonies of cells transformed with mutagenized sequences corresponding to differentially regulated genes that produce an increased amount of hydrogen under a given set of conditions than the starting strain or cells transformed with only vector and/or marker sequences are identified in this screening step. These novel strains are then utilized for the production of hydrogen.

[116] In one embodiment, the DNA construct, or substantial parts of the DNA construct, containing the mutagenized sequences is cloned, amplified, or otherwise recovered from a first strain that generates an increased amount of hydrogen. The DNA construct is put through further mutagenesis protocols to generate a new library of DNA constructs used for further screening or selection of new strains that generate increased amounts of hydrogen compared to the originally identified first strain.

[117] Nucleic acid constructs used for transforming cells may be in circular form or linear form. In addition, such constructs may be comprised of DNA or RNA. For instance, bacterial artificial chromosomes may utilized and are comprised of DNA. Alternatively, RNA vectors, such as viruses, may also be used. Viral transformation protocols for microbes are well known in the art.

[118] In one embodiment, cells are screened for increased production of hydrogen in a high-throughput fashion after being grown on solid culture media. Colonies are identified as novel strains that produce increased amounts of hydrogen. The mutagenized sequences that impart the phenotype of the ability to produce increased amounts of hydrogen are isolated from each strain of the plurality of colonies. The isolated sequences are then put through another round of shuffling, in which the sequences are randomly cleaved, denatured, reannealed, and extended using a polymerase to generate a new library of mutagenized sequences. The sequences are then used to transform strains of the host microbe in a new round of screening or selection to generate further novel strains that produce increased amounts of hydrogen compared to the previous plurality of colonies. This process is repeated as many times as desired. High throughput methods of manipulating cells are well known in the art, and cells can be plated on solid media in densities of 9 colonies or more per square inch (Hicks, Plant Physiol 2001 Dec;127(4):1334-8).

Mating of Strains

[119] In one embodiment, different differentially regulated genes are mutagenized and used to transform cells for screening or selection for transformants that generate an increased amount of hydrogen. Transformants that have been transformed with mutagenized nucleic acid sequences corresponding to different differentially regulated genes are then mated to each other to provide progeny containing different combinations of mutagenized nucleic acid sequences. The progeny are then screened or selected for the ability to generate an increased amount of hydrogen. Screenable or selectable markers may be excised through such techniques as the Cre-lox system or FLP

recombinase. Mating protocols, such as protoplast fusion, are known in the art. In addition, mating protocols for organisms such as green algae are also known (Harris, (1989) The Chlamydomonas Sourcebook. Academic Press, New York).

[120] In another embodiment, cells that produce an increased amount of hydrogen due to random mutagenesis, such as chemical or insertion mutagenesis, are mated to cells that produce an increased amount of hydrogen due to mutagenized nucleic acid sequences corresponding to genes that are involved in a hydrogen production pathway. The progeny from the mating are screened or selected for the ability to generate an increased amount of hydrogen compared to any parental strain. Any strain that differs in genome sequence from a wild-type strain that produces an increased amount of hydrogen compared to the strain from which it is derived can be mated to a second strain distinct in genome sequence from the first strain that also produces an increased amount of hydrogen compared to the strain from which it is derived. Progeny from the mating are screened or selected for the ability to produce an increased amount of hydrogen compared to either parent. This type of mating, referred to as pairwise mating, is depicted in figure 11.

[121] In another embodiment, three or more strains that have distinct genome sequences and produce an increased amount of hydrogen are mated to each other in a multiparental mating reaction, and the progeny are screened or selected for the ability to produce an increased amount of hydrogen compared parental strains. In green algae multiparental mating, cells are induced to undergo gametogenesis by removing nitrogen from the media. Cells mate to form zygospores. The cells are induced to germinate by adding nitrogen back to the media. The population is then induced to mate again by removing nitrogen to induce gametogenesis again, followed by adding nitrogen back to the media. The process can be repeated as many times as desired, allowing for shuffling of genomes. Because green algae are of mating type + or -, and because cells only mate with cells of the opposite mating type, at least one strain in the multiparental mating reaction must be of opposite mating type from at least one other strain in the reaction. Multiparental mating is described further in Example 3 and is depicted in figure 12. Multiparental mating in green algae such as Chlamydomonas can be achieved through cycling the level of nitrogen in the media and allowing the different strains to mate and produce progeny. Preferably more than one nitrogen deprivation mating cycle is performed before the cells are screened or elected for a desired phenotype. Multiparental mating allows multiple advantageous genetic alterations in the genome sequence of distinct strains to be concentrated into a single genome, allowing the individual phenotypic effect of each genetic alteration to be exerted in the presence of the other phenotypic effects of the other genetic alterations. Concentrating multiple advantageous genetic alterations therefore allows for additive or synergistic effects of multiple genetic alterations to achieved. In one embodiment, the progeny of the mating are screened for the ability to generate an increased amount of hydrogen compared to all parental strains using multiwell plates containing photoautotrophic culture media, where chemochromic films are placed over the multiwell plates. A major advantage of multiparental mating is that genetic alterations that originate in cells of the same mating type can be put into the same strain through repeated nitrogen cycling in a mating reaction. Progeny from multiparental mating reactions can be screened or selected for any desired phenotype, including hydrogen production, dissolved solid transport in or out of cells, ability to survive in certain environments such as high sunlight, low sunlight, or light of a certain wavelength, or ability to survive in environments such as high salt, low salt or brackish water, the ability to bind or decompose an environmental pollutant such as PCBs, heavy metals, dioxins, and other molecules, the ability to live on a certain food source, the ability to synthesize a desired molecule, a large number of chloroplasts per cell, and any other desired phenotype.

[122] In another mating embodiment that can be performed as pairwise or multiparental mating, a library of C.

reinhardtii strains, isolated from geographically diverse regions and containing naturally occurring single nucleotide polymorphisms (SNPs), is subjected to mating and screening or selection for a desired phenotype such as hydrogen production. The strains are subjected to the above-described mating protocols, with or without mutagenesis of the strains before or after mating. In one embodiment, the cells are transformed with an expression vector constitutively expressing an iron hydrogenase before they are mated and screened or selected for the ability to generate an increased amount of hydrogen. In one embodiment, the strains that are subjected to mating are selected from the group of strains comprising (using the strain numbers of the Chlamydomonas Genetics Center, Duke University): CC-124, CC-125, CC-1690, CC-1692, CC-407, CC-408, CC-1952, CC-2290, CC-2342, CC-2343, CC-2344, CC-2931, CC-2932, CC-2935, CC-2936, CC-2937, CC-2938, CC-2935, CC-2936, CC-2937, CC-2938, CC-3059, CC-3060, CC-3061, CC-3062, CC-3063, CC-3064, CC-3065, CC-3067, CC-3068, CC-3071, CC-3073, CC-3074, CC-3075, CC-3076, CC-3078, CC-3079, CC-3080, CC-3082, CC-3083, CC-3084, CC-3086, CC-1373 and CC-3087. These strains were isolated from geographically diverse regions and contain SNPs relative to each other's genome. These strains can also be assayed for phenotypes other than hydrogen production, such as those described in the preceding paragraph.

[123] The multiparental mating can also be between cells other than Chlamydomonas, and the stimulus to induce gametogenesis can be other than nitrogen or other nutrient deprivation. For example, the stimulus can be the removal of light during exponential growth followed by addition of light in mating reactions with diatoms such as T. weissfloggi (Armbrust EV Appl Environ Microbiol. 1999 Jul;65(7):3121-8). Alternatively, the stimulus can be addition of a compound or element such as 1 mg/liter Chromium (VI) to cells such as Scenedesmus acutus (Corradi, Ecotoxicol Environ Saf. 1995 Oct;32(1):12-8; Corradi, Ecotoxicol Environ Saf. 1995 Mar;30(2):106-10.).

[124] In another embodiment, promoter sequences from a plurality of genes in the genome of an organism are used to transform cells, followed by screening or selection for a desired phenotype. For example, a plurality of 500, 1000, 1500, 2000, or more base pair promoters are amplified from the C. reinhardtii genome. The full genome sequence has been completed and can be found at http://genome.jgi-psf.org/chlre1/chlre1.home.html. The promoter sequences are connected to a selectable marker sequence and used to transform the nuclear and/or chloroplast and/or mitochondrial genome. The surviving transformants are screened or selected for a desired phenotype. Preferably, the transformants are screened for a phenotype related to a metabolic function such as the ability to produce hydrogen. Optionally, independent transformants of promoter contructs that produce an increased amount of hydrogen are mated and the progeny are screened for a further increased amount of hydrogen over any of the parents. The mating can be paiurwise or multiparental.

Methods of producing hydrogen

[125] In one embodiment, cells containing mutagenized nucleic acid sequences and capable of producing an increased amount of hydrogen are cultured in a culture container with a transparent top section in an outdoor environment. Cells are grown in minimal culture media containing water, trace amounts of metals, and inorganic salts. Preferably only photoautotrophic organisms can live in the media. Atmospheric air contacts the top surface of the culture media. Nucleic acid sequences that are involved in the production of hydrogen are transcribed from constitutive, light-induced, or dark-induced promoters. Hydrogen evolved from cells is removed from the top of the culture container. During non-daylight hours, cells, for example, become dormant, metabolize molecules such as acetate to replenish substrate for digestion and hydrogen production during daylight, or produce hydrogen through a non-photosynthetic pathway. Optionally, cells are synchronized to the same phase of the cell cycle when producing

hydrogen.

EXAMPLE 1

[126] Step 1: Sequence design: Unique sequences a-l were searched for similarity to known sequences in the Chlamydomonas genome using the WU-Blast 2.0 program on databases of the Chlamydomonas Genome Project, located at (http://www.biology.duke.edu/chlamy_genome/blast/blast_form.html). The search produced no high scoring segment pairs. The following databases were searched: Contig Set, EST clones, S1D2 ESTs, Volvocales (non-EST), and BAC-ends (JGI). Searches were performed using the WU-blastn program using the default matrix blosum62. Gapped alignments were allowed for. The default expected threshold, filter, word length, and cutoff scores were used. The sum statistics option was used for assessing the significance of aligned pairs. Primer and chimeric oligonucleotide sequences were designed using sequences from the lhcb1 gene promoter (SEQ. ID NO 1), the 3' untranslated region of the RBCS2 gene (SEQ. ID NO 3), and a selectable marker cassette (SEQ. ID NO 2).

[127] Step 2: Culturing microbes under conditions not conducive and more conducive to the generation of hydrogen: Chlamydomonas reinhardtii (strain cc-124, Chlamydomonas Genetics Center, Duke University, Durham, N.C.) is cultured under conditions not conducive to the generation of hydrogen (photoheterotrophically on Tris-acetate-phosphate medium (TAP), pH 7.2 (Harris, (1989) The Chlamydomonas Sourcebook. Academic Press, New York; Melis, Plant Physiol (2000) Jan;122(1):127-36). The culture is bubbled with 3% CO_2 in air, stirred gently (at approximately 400 rpm) at 25° C, under continuous illumination (approximately 300 $\mu E\ m^{-2}\ s^{-1}$). The cells are grown until mid-log phase (approximately 4 x 10^6 cells mL^{-1}) and then harvested by centrifugation at 2000 x g for 5 minutes. The pellet is divided half. mRNA is purified from one half of the pellet immediately after harvesting, as specified below, without freezing. The other half is washed 2 times in TAP-minus-sulfur and resuspended in the same medium to a final concentration of 4-5 x 10^6 cells mL^{-1} (Zhang, Planta (2002) Feb;214(4):552-61; Melis, Plant Physiol (2000) Jan;122(1):127-36). The cells are cultured in containers sealed from the atmosphere, under illumination (approximately 300 $\mu E\ m^{-2}\ s^{-1}$), and are gently stirred at approximately 400 rpm. The containers allow gas evolved from the algae to escape into the atmosphere but do not allow atmospheric gas to enter the culture. The cells are cultured under these conditions for approximately 60 hours. The cells are then harvested by centrifugation at 2000 x g for 5 minutes. RNA is purified immediately after harvesting, without freezing of the cell pellet.

[128] Step 3: mRNA purification: mRNA is purified from both sets of cells using the Qiagen Oligotex® system (compositions of buffers OL1, ODB, and OW1 are proprietary; these buffers are purchased directly from Qiagen Inc., Valencia, CA). DEPC-treated water is used to make all buffers. 2-5 x 10^7 cells are separated from the pellet for mRNA purification. The Oligotex® reagent is heated to 37°C in a water bath, vortexed, and set out at room temperature. 5mM Tris•Cl pH 7.5 is heated at 70°C. All supernatant is removed from cell pellets. 800 µL of 10 mM Tris•Cl pH 7.5, 140 mM NaCl, 5 mM KCl, 1% Nonidet P-40, 1 mM DTT, and (optionally with RNase inhibitors added), chilled at 4°C, is added and the pellet is resuspended. The suspension is incubated on ice for 5 minutes. The suspension is pelleted in a microcentrifuge tube for 2 minutes at between 300-500 x g at 4°C. The supernatant is transferred to a new tube. 800 µL of room temperature 1M LiCl, 20 mM Tris•Cl pH 7.5, 2 mM EDTA, 1% SDS and 145 µL of the Oligotex® suspension are added to the supernatant, which is then vortexed. The resulting mixture is then incubated at 70°C for 3 minutes and then at 20-30°C for 10 minutes. The mixture is pelleted in a microcentrifuge at 14,000-18,000 x g for 5 minutes. The supernatant is removed. The pellet is resuspended in 200 µL of Qiagen buffer OL1 (containing 14.3 µL ß–mercaptoethanol per mL of OL1). 800 µL of

Qiagen buffer ODB is added and the suspension is incubated at 70°C for 3 minutes and room temperature for 10 minutes. The suspension is then pelleted in a microcentrifuge at maximum speed for 5 minutes. The supernatant is removed. The pellet is then resuspended in 600 µL of Qiagen buffer OW1. The suspension is then pipetted onto a large Qiagen Oligotex spin column placed inside a 2 mL microcentrifuge tube and is centrifuged for 1 minute at maximum speed. The spin column is then placed in an RNase-free 2 mL microcentrifuge tube. 600 µL of 10 mM Tris•Cl pH 7.5, 1 mM EDTA, 150 mM NaCl is added to the spin column, which is then centrifuged for 1 minute at maximum speed. The flow through is discarded and 600 µL of 10 mM Tris•Cl pH 7.5, 1 mM EDTA, 150 mM NaCl is added to the spin column, which is then centrifuged again for 1 minute at maximum speed. The spin column is then placed in a new Rnase-free 2 mL microcentrifuge tube. Approximately 200 µL of 70°C 5 mM Tris•Cl pH 7.5 is added to the spin column. The resin is resuspended by pipetting the buffer:resin mix several times. The spin column is then centrifuged for 1 minute at maximum speed. The flow through is pipetted to a new RNase-free tube. The elution process is repeated with another 200 µL of 70°C 5 mM Tris•Cl pH 7.5 and the flow through is added to the first flow through. The concentration and purity of the RNA is analyzed using spectrophotometric analysis.

[129] Step 4: cDNA synthesis and in vitro transcription: Double stranded, labeled, cDNA is synthesized from the purified mRNA samples using the Invitrogen Life Technologies Superscript® Choice system (Invitrogen Inc., Carlsbad, Ca.). mRNA samples from cells cultured under conditions not conducive to the generation of hydrogen and from cells cultured under conditions more conducive to the generation of hydrogen are processed simultaneously. 4 µg of mRNA from each sample are put into RNAse-free microcentrifuge tubes, along with 100 pmol HPLC-purified primer of the sequence 5'-GGCCAGTGAATTGTAATACGACTCACTATAGGGAGGCGG-$(dT)_{24}$-3'. The tube is incubated at 70°C for 10 minutes, briefly centrifuged, and placed on ice for 5 minutes. The following reagents are added: (1) 1 µL 10 mM dNTP mix; (2) 2 µL 100 mM DTT; (3) 4 µL 5X first strand cDNA buffer (proprietary composition, available from Invitrogen Inc, Carlsbad, Ca.). The reaction is then incubated at 37°C for 2 minutes. 4 µL of 200U/µL SuperScript® II reverse transcriptase is added to the reaction to make a final volume of 20 µL. The reaction is then incubated at 37°C for 1 hour. The reaction is then placed on ice and the following regents are added and mixed: 91 µL of DEPC-treated water, 30 µL of 5X second strand reaction buffer (proprietary composition, available from Invitrogen Inc, Carsbad, Ca.), 3 µL of 10 mM dNTP mix, 1 µL of 10 U/µL E. coli DNA ligase, 4 µL of 10 U/µL E. coli DNA polymerase I, and 1 µL of 2 U/µL E. coli Rnase H. The reaction is incubated at 16°C for 2 hours. 2 µL of 5U/µL T4 DNA Polymerase is added to the reaction and it is incubated for 5 minutes at 16°C. 10 µL 0.5M EDTA is added to the reaction.

[130] The reaction is put through a phenol:chloroform extraction using a Phase-Lock gel (optionally the reaction is put through a standard phenol:chloroform extraction). The Phase-Lock gel is pelleted in a 1.5 mL microcentrifuge tube at 12,000 x g for 30 seconds. 162 µL of 25:24:1 phenol:chloroform:isoamyl alcohol (saturated with 10 mM Tris-HCl pH 8.0, 1 mM EDTA) is added to the 162 µL reaction to a total 324 µL. The mixture is briefly vortexed, and the entire 324 µL is then added to the Phase-Lock gel tube. The tube is centrifuged at ≥12,000 x g for 2 minutes. The upper aqueous layer containing the cDNAs is transferred to a new 1.5 mL tube. 0.5 volumes of 7.5 M NH_4OAc and 2.5 volumes of 100% ethanol are added to the cDNAs. The tube is vortexed and then centrifuged at ≥12,000 x g for 20 minutes. The supernatant is removed and the pellet is washed with 500 µL of 80% ethanol. The tube is then centrifuged at ≥12,000 x g for 5 minutes. The wash is repeated once. The pellet is then air dried and resuspended in 12 µL RNase-free water. The cDNA sample from cells cultured under conditions conducive to the

generation of hydrogen is labeled as the "conducive C. rein sample." The cDNA sample from cells cultured under conditions not conducive to the generation of hydrogen is labeled as the "nonconducive C. rein sample." The cDNA samples are put through in vitro transcription reactions and are biotin labeled using the Enzo® BioArray® High Yield RNA Labeling Kit (available as part No. 900182 from Affymetrix Inc. Santa Clara, CA).

[131] Step 5: Labeled in vitro transcript purification: Total amounts of RNA generated from the in vitro transcription reactions are determined by spectrophotometric and/or gel electrophoresis. Biotin-labeled RNA samples that originated from cells cultured under conditions not conducive to the generation of hydrogen and biotin-labeled RNA samples that originated from cells cultured under conditions more conducive to the generation of hydrogen are processed simultaneously. 600-800 μg of biotin-labeled RNA are purified on Qiagen RNeasy® midi columns. All centrifugations and reactions are performed at room temperature. For smaller or larger amounts of biotin-labeled RNA, mini or maxi columns are used, respectively, along with modified protocols according to the manufacturer. The labeled RNA is added to a tube, and is brought up to a volume of 1 mL with RNAse-free water. 4 mL of buffer RLT is added (compositions of buffers RLT, RW1, and RPE are proprietary; these buffers are purchased directly from Qiagen Inc., Valencia, CA) and the sample is mixed. 2.8 mL 100% ethanol and the sample is mixed. The sample is immediately applied to a Qiagen RNeasy® midi column, which is placed in a 50 mL tube, and centrifuged 5 minutes at 3,000-5,000 x g. The flow through is discarded. 2.5 mL of buffer RPE is added to the column, which is then centrifuged 2 minutes at 3,000-5,000 x g. The flow through is discarded. 2.5 mL of buffer RPE is again added to the column, which is then centrifuged 5 minutes at 3,000-5,000 x g. The column is placed in a new 15 mL RNase-free tube. 250 μL of RNase-free water is added to the column. The column is allowed to sit for 1 minute and is then centrifuged 3 minutes at 3,000-5,000 x g. Another 250 μL of RNase-free water is added to the column. The column is allowed to sit for 1 minute and is then centrifuged 3 minutes at 3,000-5,000 x g. The concentration of the eluted biotin-labeled RNA is measured spectrophotometrically. If the concentration is less than 0.6 μg/μL, the biotin-labeled RNA is precipitated by adding 0.5 volumes 7.5 M NH_4OAc and 2.5 volumes 100% ethanol and resuspended in a smaller volume of RNase free water. The tube is vortexed and then placed at –20°C for at least 1 hour. The tube is centrifuged at ≥12,000 x g at 4°C for 30 minutes. The pellet is washed twice with 500 μL of –20°C 80% ethanol. The pellet is air dried and resuspended in 10 μL RNase-free water. The concentration of biotin-labeled RNA is adjusted to 2 μg/μL.

[132] Step 6: Labeled in vitro transcript fragmentation :12 μL of 2 μg/μL biotin-labeled RNA is added to an RNase-free tube along with 3 μL of 5X fragmentation buffer (200 mM Tris-acetate pH 8.1, 500 mM KOAc, 150 mM MgOAc). The tube is placed at 94°C for 35 minutes and then placed on ice. The biotin-labeled RNA is fragmented into sizes from approximately 35-200 nucleotides, and this is confirmed by gel electrophoresis using appropriate size markers.

[133] Step 7: Microarray hybridization and differential expression profile creation: Microarray chips containing 2,761 unique C. reinhardtii sequences are obtained from the Chlamydomonas Genome Project (Duke University, Durham, N.C. http://www.biology.duke.edu/chlamy_genome/microarrays.html). Sequence IDs and grid locations for clones are obtained from the same source (at ftp://ftp.biology.duke.edu/pub/chlamy_genome/sequences/). Fragmented biotin labeled RNA samples are hybridized to C. reinhardtii microarrays according to Affymetrix GeneChip Expression Analysis protocols (Affymetrix Inc., Santa Clara, CA.). Microarrays with labeled nonconductive RNA samples hybridized and microarrays with labeled conducive RNA samples hybridized are compared and analyzed for identification of differentially regulated genes. The microarray data set containing the expression data from cells cultured under

conditions not conducive to the generation of hydrogen and cells cultured under conditions more conducive to the generation of hydrogen is a differential expression profile.

[134] Step 8: Creation of probes corresponding to differentially regulated genes: Genes that exhibit greater than a 1.5-fold difference in expression between cells cultured under conditions not conducive to the generation of hydrogen and cells cultured under conditions more conducive to the generation of hydrogen are identified as differentially regulated genes. The 5 genes (referred to hereinafter as the 1 H_2, 2 H_2, 3H_2, 4 H_2, and 5 H_2 genes, and collectively as the 1-5 H_2 set) that are not expressed in cells cultured under conditions not conducive to the generation of hydrogen and are upregulated most compared to other upregulated genes when cells are switched from conditions not conducive to the generation of hydrogen to conditions more conducive to the generation of hydrogen are selected for mutagenesis. Alternatively, the iron-hydrogenase gene is designated as on of the 5 genes, regardless of its expression level relative to other genes. PCR primers are designed corresponding to a 50-200 base pair segment of each gene of the 1-5 H_2 set, wherein the segment chosen does not contain a specific restriction enzyme site corresponding to restriction enzymes that leave 5' overhangs at cut sites. For example, the restriction enzymes BamHI, Hind III, and Bgl II leave 5' overhangs after cutting double stranded DNA. The PCR primers contain the restriction enzyme sequence chosen at their 5' end. The primers are used to amplify their corresponding fragment from each gene of the 1-5 H_2 set using the conducive C. rein cDNA sample as a template. PCR products are digested with the restriction enzyme corresponding to the ends of amplified fragments. The PCR products are purified from the digested ends using agarose gel electrophoresis and electroelution from the gel fragment. The electroeluted PCR products, referred to hereinafter as the 1-5 H_2 set probes, are precipitated from the electroelution buffer with 0.5 volumes of 7.5 M NH_4OAc and 2 volumes of –20°C 100% ethanol. The 1-5 H_2 set probes are pelleted at 14,000 x g. The pellets are washed two times with –20°C 70% ethanol. The pellets are dried and resuspended in water.

[135] Step 9: Culturing microbes capable of producing hydrogen and creation of cDNA libraries: The following species of Chlamydomonas are cultured under conditions more conducive to the generation of hydrogen (available from the UTEX collection at The University of Texas at Austin, Austin, TX): (1) Chlamydomonas pulvinata (UTEX strain number 212, isolated from Switzerland); (2) Chlamydomonas pygmaea (UTEX strain number 2539, isolated from Prudhoe Bay, Alaska); (3) Chlamydomonas radiata (UTEX strain number 966, isolated from McMahan, Texas); (4) Chlamydomonas rapa (UTEX strain number 1342, isolated from Danube River, Bratislava, Czechoslovakia); (5) Chlamydomonas sajao (UTEX strain number 2277, isolated from Sa Jiao, China); (6) Chlamydomonas segnis^{222} (UTEX strain number 222, isolated from West Humble, Surrey, England); (7) Chlamydomonas segnis^{1638} (UTEX strain number 1638, isolated from Dauphin Is., Alabama, U.S.A.); (8) Chlamydomonas segnis^{1919} (UTEX strain number 1919, isolated from Delta Marsh; Manitoba, Canada); (9) Chlamydomonas smithii (UTEX strain number 1061, isolated from Santa Cruz, California, U.S.A.); (10) Chlamydomonas sphaeroides (UTEX strain number 221, isolated from India); (11) Chlamydomonas surtseyiensis (UTEX strain number 1796, isolated from Surtsey, Iceland); (12) Chlamydomonas ulvaensis (UTEX strain number 724, isolated from Ulva Island, Scotland); (13) Chlamydomonas zimbabwiensis (UTEX strain number 2213, isolated from Zimbabwe); (14) Chlamydomonas reinhardtii (strain cc124, Chlamydomonas Genetics Center, Duke University, Durham, N.C.). The species are cultured in TAP-minus-sulfur medium. The cells are cultured in containers sealed from the atmosphere, under illumination (approximately 300 uE m^{-2} s^{-1}), and are gently stirred at approximately 400 rpm. The containers allow gas evolved from the algae to escape into the atmosphere but do not allow atmospheric gas to enter the culture. The cells are cultured under these conditions for approximately 60

hours. The cells are then harvested by centrifugation at 2000 x g for 5 minutes. mRNA is purified immediately after harvesting, without freezing of the cell pellets. mRNA is purified from each Chlamydomonas strain as previously described using the Qiagen Oligotex® system.

[136] cDNA libraries are made from each Chlamydomonas mRNA sample. Double stranded cDNA is synthesized from the purified mRNA samples using the Invitrogen Life Technologies Superscript® Choice system. mRNA samples from each Chlamydomonas strain are processed in parallel. 4 μL of 1 μg/μL mRNA in DEPC-treated water is added to an RNase-free centrifuge tube. 2 μL of 0.5 μg/μL oligo$(dT)_{12-18}$ primer and 2 μL of 50ng/μL of random hexamer primers are added to the mRNA. The sample is heated at 70°C for 10 minutes and immediately transferred to ice. The sample is briefly centrifuged and the following components are added: (1) 4 μL of 250 mM Tris•HCl pH 8.3, 375 mM KCl, 15 mM $MgCl_2$; (2) 2 μL of 100 mM DTT; (3) 1 μL of 10mM dNTPs; (4) 1 μL 1 μCi/μL [α-^{32}P]dCTP. The reaction is mixed and incubated at 37°C for 2 minutes. 4 μL of 200 U/μL of SuperScript® Reverse Transcriptase II is added to the reaction, which is mixed and incubated at 37°C for one hour and then placed on ice. 18 μL of the reaction is placed into a new tube. The following reagents are also added: (1) 93 μL of DEPC-treated water; (2) 30 μL of 100 mM Tris•HCl pH 6.9, 450 mM KCl, 23 mM $MgCl_2$, 0.75 mM β–NAD^+, 50 mM $(NH_4)_2SO_4$; (3) 3 μL 10 mM dNTPs; (4) 1 μL of 10 U/μL E. coli DNA ligase; (5) 4 μL of 10 U/μL E. coli DNA Polymerase I; (6) 1 μL of 2 U/μL E. coli RNase H. The reaction is briefly vortexed, briefly centrifuged, and incubated for 2 hours at 16°C. 2 μL of 5 U/μL T4 DNA Polymerase is added and the reaction is incubated 5 minutes at 16°C. The reaction is then placed on ice and 10 μL of 0.5 M EDTA is added. 150 μL of 25:24:1 phenol:chloroform:isoamyl alcohol is added to the reaction, which is then vortexed and centrifuged at room temperature for 5 minutes at 14,000 x g. 140 μL of the upper aqueous phase is transferred to a new microcentrifuge tube. 70 μL of 7.5 M NH_4OAc and 500 μL of -20°C 100% ethanol are added to the sample. The tube is vortexed and centrifuged at room temperature for 5 minutes at 14,000 x g. The supernatant is removed and the pellet is washed with 500 μL of –20°C 70% ethanol. The tube is centrifuged at room temperature for 2 minutes at 14,000 x g and the supernatant is discarded. The pellet is dried at 37°C for 10 minutes. The pellet is resuspended in: (1) 18 μL of DEPC-treated water; (2) 10 μL of 330 mM Tris•HCl pH 7.6, 50 mM $MgCl_2$, 5 mM ATP; (3) 10 μL of 1 μg/μL EcoRI (Not I) adapters; (4) 7 μL of 100 mM DTT; (5) 5 μL of 1 U/μL T4 DNA ligase. The reaction is mixed and incubated for 24 hours at 16°C. The reaction is then incubated at 70°C for 10 minutes and then placed on ice. 3 μL of 10 U/μL T4 Polynucleotide Kinase is added to the sample, which is mixed and then incubated for 0.5 hours at 37°C. The reaction is then incubated for 10 minutes at 70°C and placed on ice. For each sample, a 1 mL pre-packed Sephacryl S-500 HR column is drained of 20% ethanol. 800 μL of 10 mM Tris•HCl pH 7.5, 0.1 mM EDTA, 25 mM NaCl is pipetted onto the top of each column. The column is allowed to drain. The wash is performed 3 more times with the same volume. 97 μL of 10 mM Tris•HCl pH 7.5, 0.1 mM EDTA, 25 mM NaCl is added to each reaction and mixed. The reaction is added to the top of the tube and drained into a first microcentrifuge tube. 100 μL of 10 mM Tris•HCl pH 7.5, 0.1 mM EDTA, 25 mM NaCl is added to the top of the column and drained into a second microcentrifuge tube. 100 μL of 10 mM Tris•HCl pH 7.5, 0.1 mM EDTA, 25 mM NaCl is added to the top of the column and each drop flowing from the bottom of the tube is collected into a new tube. The process is continued with 100 μL of 10 mM Tris•HCl pH 7.5, 0.1 mM EDTA, 25 mM NaCl being added to the top of the column until 18 drops are collected in 18 successive tubes numbered 3-20. The volume in all 20 tubes is measured. The numerical volume of each tube is added to determine the fraction of column flow through in each tube. Tubes containing volume collected after 600 μL of eluate has flowed through the column are discarded. The remaining tubes are placed in a scintillation counter and Cerenkov counts for each tube are

measured. Tubes containing only background Cerenkov counts are discarded. The concentration of cDNA in each remaining fraction is determined according to the SuperScript® Choice System for cDNA Synthesis manufacturer's recommendations (Invitrogen Inc., Carlsbad, Ca., Catalog Series 18090). Fractions containing more than 0.1 ng/μL cDNA are pooled. The cDNAs are precipitated with 0.5 volumes of 7.5 M NH_4OAc and 2 volumes of –20°C 100% ethanol. The sample is vortexed and centrifuged at room temperature for 20 minutes at 14,000 x g. The pellet is washed two times with 500 μL of –20°C 70% ethanol and then dried at 37°C for 10 minutes. The pellet is resuspended in 20 μL 10 mM Tris•HCl pH 7.5, 0.1 mM EDTA, 25 mM NaCl. A dilution of each Chlamydomonas cDNA is made to yield 10 μL of 1 ng/μL cDNA in 10 mM Tris•HCl pH 7.5, 0.1 mM EDTA, 25 mM NaCl. All Chlamydomonas cDNA samples are processed in parallel. To each cDNA tube, the following reagents are added: (1) 4 μL of 250 mM Tris•HCl pH 7.6, 50 mM MgCl2, 5 mM ATP, 5 mM DTT, 25% (w/v) Polyethylene glycol 8000; (2) 5 μL of 10 ng/μL, EcoRI cut, dephosphorylated plasmid pcDNA3(+) (available from Invitrogen Inc., Carlsbad, Ca.); (3) 1 μL of 1 U/μL T4 DNA ligase. The reaction, hereinafter referred to for each strain as the "X strain conducive cDNA library" (such as the Chlamydomonas surtseyiensis conducive cDNA library), is incubated 3 hours at room temperature and then frozen at –20°C.

[137] Step 10: Cloning of 1-5 H_2 set cDNAs: The 1-5 H_2 set probes are labeled with [α-^{32}P]dNTPs using the Klenow DNA Polymerase fragment (available from New England Biolabs Inc., Beverly, MA) according to standard protocols. The conducive cDNA libraries from the fourteen Chlamydomonas strains grown in step 9 are used to transform competent E. coli cells using standard protocols. The plated E. coli cells transformed with each of the fourteen conducive cDNA libraries are used for cloning cDNAs for each of the 1-5 H_2 set gene homologues from each of the fourteen conducive cDNA libraries using standard cDNA cloning methods (Maniatis et al. (1989) Molecular Cloning : A Laboratory Manual Cold Spring Harbor Laboratory). The probes used to identify each of the 1-5 H_2 set gene homologues are the 1-5 H_2 set probes. The identified clones are sequenced. Full length cDNAs are obtained using RACE-PCR with mRNA samples from each Chlamydomonas strain as template. A full length cDNA from each of the 1-5 H_2 set gene homologues is selected for use in DNA shuffling and is referred to as the X strain Y H_2 gene (such as the Chlamydomonas pygmaea 3 H_2 gene). A total of 70 cDNA sequences are obtained (a 1 H_2, 2 H_2, 3 H_2, 4 H_2, and 5 H_2 gene from each of the 14 Chlamydomonas strains).

[138] Step 11: Creation of nonshuffled DNA construct segments: Nonshuffled segments I-VIII are generated through PCR amplification using primers and templates listed in Table 1. The position of these primers relative to the sequence information they contain (not drawn to scale) is depicted in Figure 6 by arrows. Nonshuffled segments I-VIII are gel purified, electroeluted, and precipitated. The fragments are resuspended in water.

[139] Step 12: Shuffling of 1-5 H_2 set coding regions : The coding region of each of the 70 1-5 H_2 set homologue genes is amplified using the cDNA plasmid as template and primers corresponding to the N and complement of the C terminal portions of the cDNA coding sequences. PCR products corresponding to the coding regions of all 1-5 H_2 set homologue genes are gel-purified, electroeluted, precipitated, and resuspended in 50 mM Tris•HCl pH 7.4, 1 mM $MgCl_2$. Alternatively PCR primers are removed from the reaction using the Wizard® PCR product (Promega Corp, Madison, WI) and the PCR products are resuspended in 50 mM Tris•HCl pH 7.4, 1 mM $MgCl_2$. Chimeric oligonucleotides are synthesized according to Table 2 and are resuspended in 50 mM Tris•HCl pH 7.4, 1 mM $MgCl_2$.

[140] 70 PCR products corresponding to the coding regions of all 1-5 H_2 set homologue genes are quantified with spectrophotometry. Reactions for each of the 1-5 H_2 genes are performed in parallel. Equal molar amounts of each cDNA corresponding to each of the 1-5 H_2 set homologue genes are pooled in separate tubes to obtain a total of 4

ug DNA in 100 μL 50 mM Tris•HCl pH 7.4, 1 mM $MgCl_2$. In other words, 0.2857 μg of cDNA from each of the 14 cDNAs corresponding to the 1 H_2 gene are added to a single tube. 0.2857 μg of cDNA from each of the 14 cDNAs corresponding to the 2 H_2 gene are added to a different tube, and so on, such that each H_2 gene is shuffled in a separate reaction. DNAse I (obtained from Sigma Corp., St. Louis, MO) is added to each tube at a concentration of 0.0015 units of Dnase I per μl of DNA. The digestion reaction proceeds for 15 minutes at room temperature and is stopped. Digestion products from approximately 20-150 base pairs are purified from 2% low melting agarose gels, electroeluted, and precipitated. An equivalent molar amount of corresponding chimeric oligonucleotides to the original starting material for each cDNA is added to each tube. For instance, a 900 base pair 1 H_2 cDNA from one of the 14 strains corresponds to 0.481 pmol (1/14 of 4 μg added to DNAse I digestion reaction converted to pmol for a 900 base pair double stranded fragment). For 1 H_2 cDNAs of approximately 900 base pairs, 0.481 pmol of chimeric oligonucleotides 1.1-1.14 and 0.481 pmol of chimeric oligonucleotides 2.1-2.14 are added to the purified fragmented coding regions. Chimeric oligonucleotides 3.1-3.14 and 4.1-4.14 are added to 2 H_2 fragments. Chimeric oligonucleotides 5.1-5.14 and 6.1-6.14 are added to 3 H_2 fragments. Chimeric oligonucleotides 7.1-7.14 and 8.1-8.14 are added to 4 H_2 fragments. Chimeric oligonucleotides 9.1-9.14 and 10.1-10.14 are added to 5 H_2 fragments. Chimeric oligonucleotides and 20-150 base pair cDNA fragments are resuspended in 0.2 mM of each dNTP, 2.2 mM $MgCl_2$, 50 mM KCl, 10 mM Tris•HCl pH 9.0, 0.1% Triton X-100, to a volume of 100 μl where the DNA concentration is approximately 20 ng/μl. 1.25 units of Taq polymerase and 1.25 units of Pfu polymerase are added. Each of the 5 tubes corresponding to cDNA fragments and chimeric oligonucleotides for genes 1-5 H_2 are subjected to a themocycling program of 94°C for 60 seconds one time, followed by 40 cycles of 94°C for 30 seconds, 55°C for 30 seconds, and 72°C for 30 seconds, followed by a one time incubation of 72°C for 5 minutes. 10 μl from each reaction is brought up to 100 μl in new PCR tubes in 0.2 mM of each dNTP, 2.2 mM $MgCl_2$, 50 mM KCl, 10 mM Tris•HCl pH 9.0, 0.1% Triton X-100, 8 μM of primers corresponding to unique sequences and the complements of unique sequences at the ends of each cDNA fragment, and 1.25 units of Taq polymerase and 1.25 units of Pfu polymerase. Shuffled 1 H_2 genes are amplified by primers corresponding to unique sequence a and the complement of unique sequence b. Shuffled 2 H_2 genes are amplified by primers corresponding to unique sequence c and the complement of unique sequence d. Shuffled 3 H_2 genes are amplified by primers corresponding to unique sequence e and the complement of unique sequence f. Shuffled 4 H_2 genes are amplified by primers corresponding to unique sequence g and the complement of unique sequence h. Shuffled 5 H_2 genes are amplified by primers corresponding to unique sequence i and the complement of unique sequence j. The amplification reactions are performed in a thermocycler for a program of 94°C for 60 seconds one time, followed by 20 cycles of 94°C for 30 seconds, 55°C for 30 seconds, and 72°C for 30 seconds, followed by a one time incubation of 72°C for 5 minutes. PCR products, now referred to as the 1 H_2 shuffled library, the 2 H_2 shuffled library, etc., are gel purified, electroeluted, precipitated, and resuspended in water.

[141] Step 13: Synthesis of test constructs: Equimolar amounts of nonshuffled segments I-VIII and 1-5 H_2 shuffled libraries are added together in a new primerless PCR reaction. 1 pmol each of nonshuffled segment I, nonshuffled segment II, nonshuffled segment III, nonshuffled segment IV, nonshuffled segment V, nonshuffled segment VI, nonshuffled segment VII, nonshuffled segment VIII, 1 H_2 shuffled library, 2 H_2 shuffled library, 3 H_2 shuffled library, 4 H_2 shuffled library, and 5 H_2 shuffled library are brought up to a volume of 100 μl in 0.2 mM of each dNTP, 2.2 mM $MgCl_2$, 50 mM KCl, 10 mM Tris•HCl pH 9.0, 0.1% Triton X-100, with 2.5 units of Pfu DNA polymerase. The reaction is subjected to a themocycling program of 94°C for 60 seconds one time, followed by 40 cycles of 94°C for 30 seconds, 55°C for 30 seconds, and 72°C for 30 seconds, followed by a one time incubation of

72°C for 5 minutes. Double stranded primerless PCR products, now referred to as 1-5 H_2 test constructs, are separated from oligonucleotides and fragments by gel electrophoresis and products of the expected size are electroeluted, precipitated, and resuspended in sterile water.

[142] Step 14: Transformation of cells with mutagenized nucleic acid sequences: The Chlamydomonas reinhardtii strain CC-400 (a cell wall deficient strain, Chlamydomonas Genetics Center, Duke University) is grown with shaking in TAP media (Harris, (1989) The Chlamydomonas Sourcebook. Academic Press, New York; Gorman, Proc Natl Acad Sci U S A (1965) Dec;54(6):1665-9) until the cells reach a density of approximately 2 x 10^6 cells/ml. The cells are pelleted at 4000 x g for 5 minutes and the supernatant is removed. The cell pellet is resuspended in 7.5 ml per liter of original culture of TAP medium. The following components are added, in order, to 25 sterile tubes: 300 µl of cells, 1 µg of 1-5 H_2 test construct, 100 µl of sterile-filtered 20% PEG, 300 mg of sterile glass beads (prepared according to Kindle, Meth Enzymology (1998) 297: 27-38). Each tube is vortexed 15-30 seconds at high speed. The cells are removed from the tube and spread onto plates containing phleomycin (Stevens, Mol Gen Genet (1996) Apr 24;251(1):23-30). Plates are incubated in low light (approximately 5 µE m^{-2} s^{-2}) at 25°C for 4-6 days in atmospheric air until colonies appear.

[143] Step 15: Screening for increased amounts of hydrogen: Phleomycin resistant colonies are transferred to new plates containing identical culture media. Colonies are plated in 96-colony grids. Replica plates are also made and stored at 15°C in low light. The 96-colony plates, made of clear plastic, are incubated in low light (approximately 5 µE m^{-2} s^{-2}) at 25°C in atmospheric air for until colonies are approximately 3 mm in diameter. Chlamydomonas reinhardtii strain CC-400 is used as a control on each 96-colony plate. After colonies have grown to the desired size, 3 mm thick filter paper is placed over the plate, covering the colonies. A chemochromic film containing tungsten trioxide is placed on top of the filter paper (Seibert). A rectangular clear plastic grid design is placed directly over the chemochromic film such that the center of each square on the grid is directly over the center of a cell colony. The plates are incubated in light (approximately 55 µE m^{-2} s^{-2}) at 25°C in 5% oxygen for 12 hours. The plates are illuminated from above and below. After 12 hours, each plate is photographed from the top using a digital camera within 5 seconds of removal from the incubation chamber. The images are scanned by densitometry and are subsequently screened for dark spots on the chemochromic film that indicate the production of hydrogen. Spots that are quantitatively darker than spots directly over control colonies of nontransformed Chlamydomonas reinhardtii strain CC-400 indicate cells that generate an increased amount of hydrogen. These colonies are recovered from the test plates or the replica plates.

EXAMPLE 2

[144] Step 1: Sequence design: Unique sequences a-h were searched for similarity to known sequences in the Chlamydomonas genome using the WU-Blast 2.0 program on databases of the Chlamydomonas Genome Project, located at (http://www.biology.duke.edu/chlamy_genome/blast/blast_form.html). The search produced no high scoring segment pairs. The following databases were searched: Contig Set, EST clones, S1D2 ESTs, Volvocales (non-EST), and BAC-ends (JGI). Searches were performed using the WU-blastn program using the default matrix blosum62. Gapped alignments were allowed for. The default expected threshold, filter, word length, and cutoff scores were used. The sum statistics option was used for assessing the significance of aligned pairs. Primer and chimeric oligonucleotide sequences were designed using sequences from the lhcb1 gene promoter (SEQ ID 148), the 3' untranslated region of the RBCS2 gene (SEQ ID 150), and a green fluorescent protein gene (SEQ ID 179).

[145] Step 2: Obtaining cDNA sequences : cDNA sequences are obtained, using methods previously disclosed, for:

Chlamydomonas reinhardtii ferredoxin (Genbank accession number L10349, SEQ ID NO 172); Chlamydomonas reinhardtii hydrogenase (Genbank accession number AF289201, SEQ ID NO 173); Scenedesmus obliquus hydrogenase (Genbank accession number AJ271546, SEQ ID NO 177), and Chlorella fusca hydrogenase (Genbank accession number AJ298227, SEQ ID NO 178). cDNA sequences are identified using synthetic oligonucleotides corresponding to GenBank sequences as probes.

[146] The coding region of each of the 3 iron hydrogenase genes is amplified using the cDNA plasmid as template and primers corresponding to the N and complement of the C terminal portions of the coding regions of the cDNA sequences. PCR products corresponding to the coding regions of the 6 hydrogenase genes are gel-purified, electroeluted, precipitated and resuspended in 50 mM Tris•HCl pH 7.4, 1 mM $MgCl_2$. Alternatively PCR primers are removed from the reaction using the Wizard® PCR product and the PCR products are resuspended in 50 mM Tris•HCl pH 7.4, 1 mM $MgCl_2$. Chimeric oligonucleotides are synthesized according to Table 4 and are resuspended in 50 mM Tris•HCl pH 7.4, 1 mM $MgCl_2$.

[147] Step 3: Shuffling of hydrogenase coding regions: PCR products corresponding to the coding regions of the 6 hydrogenase genes are quantified using spectrophotometry. Equal molar amounts of each PCR product are pooled to obtain a total of 4 ug DNA in 100 µL 50 mM Tris•HCl pH 7.4, 1 mM $MgCl_2$. DNAse I is added at a concentration of 0.15 units of Dnase I per 100 µl of reaction volume. The digestion reaction proceeds for 15 minutes at room temperature and is stopped. Digestion products from approximately 20-150 base pairs are purified from 2% low melting agarose gels, electroeluted, precipitated, and resuspended in water. 0.7123 pmol of chimeric oligonucleotides 11.1, 11.2, 11.3, 11.4, 11.5, 11.6, 12.1, 12.2, 12.3, 12.4 12.5, and 12.6 are added to each tube. Chimeric oligonucleotides and 20-150 base pair hydrogenase coding region fragments are resuspended in 0.2 mM of each dNTP, 2.2 mM $MgCl_2$, 50 mM KCl, 10 mM Tris•HCl pH 9.0, 0.1% Triton X-100, to a volume of 100 µl where the DNA concentration is approximately 20 ng/µl. 1.25 units of Taq polymerase and 1.25 units of Pfu polymerase are added. The reaction is subjected to a themocycling program of 94°C for 60 seconds one time, followed by 40 cycles of 94°C for 30 seconds, 55°C for 30 seconds, and 72°C for 30 seconds, followed by a one time incubation of 72°C for 5 minutes. 10 µl from the reaction is brought up to 100 µl in new PCR tubes in 0.2 mM of each dNTP, 2.2 mM $MgCl_2$, 50 mM KCl, 10 mM Tris•HCl pH 9.0, 0.1% Triton X-100, 8 µM of unique sequence b and the complement of unique sequence c primers, and 1.25 units of Taq polymerase and 1.25 units of Pfu polymerase. The amplification reaction is performed in a thermocycler for a program of 94°C for 60 seconds one time, followed by 20 cycles of 94°C for 30 seconds, 55°C for 30 seconds, and 72°C for 30 seconds, followed by a one time incubation of 72°C for 5 minutes. PCR products, now referred to as the hydrogenase shuffled library, are gel purified, electroeluted, precipitated, and resuspended in water.

[148] Step 4: Error-prone PCR of ferredoxin: The Chlamydomonas reinhardtii ferredoxin coding region (SEQ ID NO 172) is amplified by PCR using primers corresponding to the N terminal and complement of the C terminal ends of the coding region. The coding region PCR product is then subjected to PCR using chimeric oligonucleotides 13 and 14. The PCR product, consisting of the Chlamydomonas reinhardtii ferredoxin coding region flanked by unique sequences d and e, is then subjected to error-prone PCR. The error-prone PCR is performed using unique sequence d and the complement of unique sequence e as primers at a concentration of 1 µM each, in a reaction also containing: 50 ng template (ferredoxin fragment flanked by unique sequences d and e), 20 mM Tris pH 8.4, 0.3 mM $MnCl_2$, 3 mM $MgCl_2$, 50 mM KCl, 0.01% gelatin, 0.2 mM dATP, 1 mM dCTP, 1 mM dGTP, 1 mM dTTP, 1 U AmpliTaq polymerase (Perkin Elmer, Foster City, CA), essentially according to the method of Leung, Technique (1989) 1, 11-15. The PCR products, now referred to as the ferredoxin library, is gel

purified, electroeluted, precipitated, and resuspended in water.

[149] Step 5: Construction of nonshuffled segments : Nonshuffled segments IX, X, XI, XII, and XIII are generated through PCR amplification using primers and templates listed in Table 3. The position of these primers relative to the sequence information they contain (not drawn to scale) is depicted in Figure 7 by arrows. Nonshuffled segments IX, X, XI, XII, and XIII are gel purified, electroeluted, and precipitated. The fragments are resuspended in water.

[150] Step 6: Construction of hydrogenase-ferredoxin test construct library: Equimolar amounts of nonshuffled segments IX, X, XI, XII, and XIII, the hydrogenase shuffled library and the ferredoxin library are added together in a new primerless PCR reaction. 1 pmol each of nonshuffled segments IX, X, XI, XII, and XIII, the hydrogenase shuffled library, and the ferredoxin library are brought up to a volume of 100 µl in 0.2 mM of each dNTP, 2.2 mM $MgCl_2$, 50 mM KCl, 10 mM Tris•HCl pH 9.0, 0.1% Triton X-100, with 2.5 units of Pfu DNA polymerase. The reaction is subjected to a themocycling program of 94°C for 60 seconds one time, followed by 40 cycles of 94°C for 30 seconds, 55°C for 30 seconds, and 72°C for 30 seconds, followed by a one time incubation of 72°C for 5 minutes. Double stranded primerless PCR products, now referred to as hydrogenase-ferredoxin test construct library, are separated from oligonucleotides and fragments by gel electrophoresis and products of the expected size are electroeluted, precipitated, and resuspended in sterile water.

[151] Step 7: Transformation of cells : The Chlamydomonas reinhardtii strain cc-400 is grown with shaking in TAP media (Harris, (1989) The Chlamydomonas Sourcebook. Academic Press, New York; Gorman , Proc Natl Acad Sci U S A (1965) Dec;54(6):1665-9) until the cells reach a density of approximately 2 x 10^6 cells/ml. The cells are pelleted at 4000 x g for 5 minutes and the supernatant is removed. The cell pellet is resuspended in 7.5 ml per liter of original culture of TAP medium. The following components are added, in order, to 25 sterile tubes: 300 µl of cells, 1 µg of hydrogenase-ferredoxin test construct, 100 µl of sterile-filtered 20% PEG, 300 mg of sterile glass beads (prepared according to Kindle, Meth Enzymology (1998) 297: 27-38). Each tube is vortexed 15-30 seconds at high speed. The cells are removed from the tube and are cultured in TAP media under continuous illumination (approximately 55 µE m^{-2} s^{-2}) at 25°C for 12 hours.

[152] Step 8: Screening cells for generation of hydrogen : Cells in media are illuminated with 395 nm light and monitored for emission at 525 nm using fluorescence-activated cell sorting (Bloodgood et al. Exp Cell Res 1987 Dec;173(2):572-85; Hegemann). Colonies exhibiting 525nm GFP emission are recovered from the sorting protocol and are plated in 96-colony grids on solid media. Replica plates are also made and stored at 15°C in low light. The 96-colony plates, made of clear plastic, are incubated in low light (approximately 5 µE m^{-2} s^{-2}) at 25°C in atmospheric air until colonies are approximately 3 mm in diameter. Chlamydomonas reinhardtii strain cc-400 is used as a control on each 96-colony plate. After colonies have grown to the desired size, 3 mm thick filter paper is placed over the plate, covering the colonies. A chemochromic film containing tungsten trioxide is placed on top of the filter paper (Seibert). A rectangular clear plastic grid design is placed directly over the chemochromic film such that the center of each square on the grid is directly over the center of a cell colony. The plates are incubated in light (approximately 55 µE m^{-2} s^{-2}) at 25°C in atmospheric air for 12 hours. The plates are illuminated from above and below. After 12 hours, each plate is photographed from the top using a digital camera within 5 seconds of removal from the incubation chamber. The images are scanned by densitometry and are subsequently screened for dark spots on the chemochromic film that indicate the production of hydrogen. Spots that are quantitatively darker than spots directly over control colonies of nontransformed Chlamydomonas reinhardtii strain cc-400 indicate cells that generate an increased amount of hydrogen. These colonies are recovered from the test plates or the replica

plates.

[153] Step 9: Isolation and further mutagenesis of hydrogenase-ferredoxin test constructs that cause increased production of hydrogen: Total DNA is isolated from the 5% of all transformant colonies exhibiting the highest level of hydrogen production. Hydrogenase-ferredoxin test constructs are recovered from the DNA by PCR using primers corresponding to unique sequence a and the complement of unique sequence h. PCR products are gel purified, electroeluted, precipitated, and resuspended in water.

[154] The hydrogenase-ferredoxin test constructs are quantified using spectrophotometry. Equimolar amounts of each recovered test construct are added to a total of 4 μg of test construct and are diluted to 100 μL to yield a reaction tube containing 50 mM Tris•HCl pH 7.4, 1 mM $MgCl_2$. DNAse I is added at a concentration of 0.15 units of Dnase I per 100 μl of reaction volume. The digestion reaction proceeds for 15 minutes at room temperature. Digestion products from approximately 20-150 base pairs are purified from 2% low melting agarose gels, electroeluted, precipitated, and resuspended in 0.2 mM of each dNTP, 2.2 mM $MgCl_2$, 50 mM KCl, 10 mM Tris•HCl pH 9.0, 0.1% Triton X-100, to a volume of 100 μl where the DNA concentration is approximately 20 ng/μl. 1.25 units of Taq polymerase and 1.25 units of Pfu polymerase are added. The reaction is subjected to a themocycling program of 94°C for 60 seconds one time, followed by 40 cycles of 94°C for 30 seconds, 55°C for 30 seconds, and 72°C for 30 seconds, followed by a one time incubation of 72°C for 5 minutes. 10 μl from the reaction is brought up to 100 μl in new PCR tubes in 0.2 mM of each dNTP, 2.2 mM $MgCl_2$, 50 mM KCl, 10 mM Tris•HCl pH 9.0, 0.1% Triton X-100, 8 μM of unique sequence a and the complement of unique sequence h primers, 1.25 units of Taq polymerase and 1.25 units of Pfu polymerase. The amplification reaction is performed in a thermocycler for with a program of 94°C for 60 seconds one time, followed by 20 cycles of 94°C for 30 seconds, 55°C for 30 seconds, and 72°C for 30 seconds, followed by a one time incubation of 72°C for 5 minutes. PCR products, now referred to as the hydrogenase-ferredoxin secondary test constructs, are gel purified, electroeluted, precipitated, and resuspended in sterile water.

[155] Step 10: Transformation of cells: The Chlamydomonas reinhardtii strain cc-400 is grown with shaking in TAP media (Harris, (1989) The Chlamydomonas Sourcebook. Academic Press, New York; Gorman , Proc Natl Acad Sci U S A (1965) Dec;54(6):1665-9) until the cells reach a density of approximately 2 x 10^6 cells/ml. The cells are pelleted at 4000 x g for 5 minutes and the supernatant is removed. The cell pellet is resuspended in 7.5 ml per liter of original culture of TAP medium. The following components are added, in order, to 25 sterile tubes: 300 μl of cells, 1 μg of hydrogenase-ferredoxin secondary test construct, 100 μl of sterile-filtered 20% PEG, 300 mg of sterile glass beads (prepared according to Kindle, Meth Enzymology (1998) 297: 27-38). Each tube is vortexed 15-30 seconds at high speed. The cells are removed from the tube and are cultured in TAP media under continuous illumination (approximately 55 μE m^{-2} s^{-2}) at 25°C for 12 hours.

[156] Step 11: Screening cells for generation of hydrogen : Cells in media are illuminated with 395 nm light and monitored for emission at 525 nm using fluorescence-activated cell sorting (Bloodgood et al. Exp Cell Res 1987 Dec;173(2):572-85; Hegemann). Colonies exhibiting 525nm GFP emission are recovered from the sorting protocol and are plated in 96-colony grids on solid media. Replica plates are also made and stored at 15°C in low light. The 96-colony plates, made of clear plastic, are incubated in low light (approximately 5 μE m^{-2} s^{-2}) at 25°C in atmospheric air until colonies are approximately 3 mm in diameter. Chlamydomonas reinhardtii strain cc-400 is used as a control on each 96-colony plate. After colonies have grown to the desired size, 3 mm thick filter paper is placed over the plate, covering the colonies. A chemochromic film containing tungsten trioxide is placed on top of the filter paper (Seibert). A rectangular clear plastic grid design is placed directly over the chemochromic film such

that the center of each square on the grid is directly over the center of a cell colony. The plates are incubated in light (approximately 55 $\mu E\ m^{-2}\ s^{-2}$) at 25°C in atmospheric air for 12 hours. The plates are illuminated from above and below. After 12 hours, each plate is photographed from the top using a digital camera within 5 seconds of removal from the incubation chamber. The images are scanned by densitometry and are subsequently screened for dark spots on the chemochromic film that indicate the production of hydrogen. Spots that are quantitatively darker than spots directly over control colonies of nontransformed Chlamydomonas reinhardtii strain cc-400 indicate cells that generate an increased amount of hydrogen. These colonies are recovered and are used for hydrogen production and/or further development.

EXAMPLE 3

Multiparental Mating Protocol

[157] 1. Place cells from 3 or more strains of algae capable of mating to each other such as Chlamydomonas reinhardtii together in the same tube, where at least one strain is of a different mating type than at least one other strain. For example, place approximately the same number of cells of the following strains into the tube: CC-124, CC-125, CC-1690, CC-1692, CC-407, CC-408, CC-1952, CC-2290, CC-2342, CC-2343, CC-2344, CC-2931, CC-2932, CC-2935, CC-2936, CC-2937, CC-2938, CC-2935, CC-2936, CC-2937, CC-2938, CC-3059, CC-3060, CC-3061, CC-3062, CC-3063, CC-3064, CC-3065, CC-3067, CC-3068, CC-3071, CC-3073, CC-3074, CC-3075, CC-3076, CC-3078, CC-3079, CC-3080, CC-3082, CC-3083, CC-3084, CC-3086, CC-1373 and CC-3087.

[158] 2. Suspend the cells nitrogen free medium, such as Sueoka's medium without NH_4Cl.

[159] 3. Incubate in light, for 12 hours, or for 1 day, or 2 days, or 3 days, or 4 days, or for 5, 6, 7, 8, 9, 10, or more days, or for fractions of the aforementioned numbers of days.

[160] 4. Add nitrogen (such as NH_4Cl) to media or move cells into nitrogen containing media and incubate in light, for 12 hours, or for 1 day, or 2 days, or 3 days, or 4 days, or for 5, 6, 7, 8, 9, 10, or more days, or for fractions of the aforementioned numbers of days.

[161] 5. Collect cells and change media back to nitrogen free and incubate in light for 12 hours, or for 1 day, or 2 days, or 3 days, or 4 days, or for 5, 6, 7, 8, 9, 10, or more days, or for fractions of the aforementioned numbers of days.

[162] 6. Repeat steps 4-5 as any times as desired.

[163] 7. Plate mating reaction on solid media (or optionally sort cells individually with a cell sorter) and pick colonies.

[164] 8. Array strains from colonies into multiwell plates containing liquid culture media.

[165] 9. Screen or select for a desired phenotype.

[166] 10. Identify 3 or more novel strains from step 9 that have the desired phenotype.

[167] 11. Repeat steps 1-9 as many times as desired.

[168] To make 1 liter of Sueoka's high salt media*:

Phosphate Buffer	50 mls
Beijerinck's stock	50 mls
Hutner's trace elements (see TAP)	1 ml
Sodium acetate	2.0 g(1.2 g if anhydrous)

[169] Phosphate Buffer

Component	For 1 liter
K_2HPO_4	28.8 g
KH_2PO_4	14.4 g

[170] Beijerinck's stock

Component	for 1 liter
NH_4Cl	10g
$MgSO_4 \cdot 7H_2O$	0.4g
$CaCl_2 \cdot 2H_2O$	0.2g

*Media for inducing gametogenesis can be made by withholding NH_4Cl from the Beijerinck's stock.

EXAMPLE 4

Gene Reassembly

[171] The process of chimeric gene assembly is depicted in figures 13-14. Sections of the active site region that are both highly conserved and correspond to the gas channel were identified using structural data, as shown in figure 9. In step 1 of figure 13, a library of approximately 110 unique Iron hydrogenase amino acid sequences was aligned using sequence manipulation software (DS Gene 1.5, Accelyrys Inc., San Diego, CA). The key in figure 15 shows the identity of amino acids from step 1 and codons from steps 2-9. In step 2, peptide sequences of conserved gas channel segments were reverse-translated into single stranded oligonucleotide sequences using *C. reinhardtii* most preferred codons from figure 10. All bars in step 1 correspond to amino acids of aligned iron hydrogenases. All bars in steps 2-9 correspond to codons that encode the amino acids from the bars of step 1. Each bar in steps 2-9 therefore depicts a codon triplet of oligonucleotide sequence. In step 3, three codons encoding amino acids that flank each side of the conserved gas channel segments were re-written to encode the corresponding *C. reinhardtii* amino acids in those flanking positions. Each oligonucleotide of step 3 therefore encodes (from left to right) three *C. reinhardtii* codons that flank the N-terminal side of a gas channel segment, followed by codons corresponding to a non-*C. reinhardtii* gas channel segment, followed by three *C. reinhardtii* codons that flank the C-terminal side of the gas channel segment. Even though these oligonucleotides encode different sequences from the *C. reinhardtii* Iron hydrogenase, the combination of recoding and the substitution of 3 flanking codons on either side of the gas channel segment generates enough nucleotide similarity that these oligonucleotides anneal to a complementary strand encoding the recoded, wild-type *C. reinhardtii* Iron hydrogenase. In step 4, the entire set of recoded oligonucleotides is mixed and annealed to single stranded "scaffold" DNA molecules that encode the wild type *C. reinhardtii* Iron hydrogenase protein in recoded form. Recoding the wild type *C. reinhardtii* iron-hydrogenase to make the scaffold achieves maximum sequence identity between the scaffold and the recoded oligonucleotides because the wild type *C. reinhardtii* Iron hydrogenase gene does not contain only the most highly preferred codons. Oligonucleotides corresponding to wild type *C. reinhardtii* gas channel segments with single residue substitutions designed to narrow the gas channel can also be mixed into in the annealing reaction. The single stranded scaffold molecule is generated by isolating the gene from a plasmid grown in a methylating host cell, followed by denaturation and separation of the strands by HPLC or other standard procedures, as described for example in U.S. patent 6,361,974. None of the primers anneal to partially overlapping sites on the C. reinhardtii strand. No

exonuclease treatment is needed to "clip" strands partially displaced by annealing of other oligonucleotide. In step 5 of figure 14, different combinations of diverse gas channel segments anneal to each full length complementary strand. Each oligonucleotide has at least 9 perfect base pairs on both ends, ensuring sufficient annealing despite internal mismatches due to sequence variation of the gas channel segments. Addition of DNA Polymerase in step 6 extends the annealed oligonucleotides, creating a combinatorial library of double stranded hybrid Iron hydrogenase molecules with numerous mismatches at "context" residue positions. Preferably the DNA Polymerase is exonuclease-deficient to prevent it from degrading parts of annealed primers in its path as it extends between annealed primers. In step 7, the methylated strands are digested using a methylation-sensitive endonuclease, as described for example in U.S. patent 6,361,974. An alternative method for separating the scaffold strands from the library strands is to use a biotinylated C-terminal primer and separate the library strands using immobilized streptavidin. In steps 8-9, an N and C terminal *C. reinhardtii* primers and DNA Polymerase are added to the library of novel Iron hydrogenase molecules for a single round of amplification. The result is a library of double stranded Iron hydrogenase sequences that have random combinations of functional gas channel segments but *C. reinhardtii* framework/hinge regions. The library is be cloned into *C. reinhardtii* cells and assayed for catalytic activity in the presence of O_2. Library members identified as active in the presence of O_2 are sequenced and a new library is made using the above method and oligonucleotides designed to anneal to a representative single stranded Iron hydrogenase identified from the first library. The screening process on the second library is performed in the presence of an additional amount of oxygen compared to the first round. This gene reassembly procdure can be used to mutagenize any nucleic acid sequence.

TABLE 1

Product	5' primer	5' primer sequence	3' primer	3' primer sequence	Template
Nonshuffled segment I	First 24 nucleotides of promoter fragment of the lhcb1 gene	5' gcagttgggtca ggggctggcgac 3'	Complement of unique sequence a-complement of last 25 base pairs of the promoter fragment of the lhcb1 gene	5'gctaagatggcc ataaggataactac ggattaacgaaatg agtctcgcccgcggc 3'	SEQ ID NO 148
Nonshuffled segment II	Unique sequence b-first 25 nucleotides of 3' UTR from RBCS2 gene	5' cgtgcatcgattaa cagcttctggacctga ccgacgtcgaccca ctctagaggat 3'	Complement of unique sequence c-complement of last 25 base pairs of the promoter fragment of the lhcb1 gene	5' cttagtcatacttg gacgtacgacgttta ataacgaaatgagt ctcgcccgcggc 3'	SEQ ID NO 151

Nonshuffled segment III	Unique sequence d-first 25 nucleotides of 3' UTR from RBCS2 gene	5' aatctgatac atgctattca gatcttacaa ccgacgtcgaccca ctctagaggat 3'	Complement of unique sequence e-complement of last 25 base pairs of the promoter fragment of the lhcb1 gene	5' agttacgatttact agtcgagtagacat tttaacgaaatgag tctcgcccgcggc 3'	SEQ ID NO 151
Nonshuffled segment IV	Unique sequence f-first 25 nucleotides of 3' UTR from RBCS2 gene	5' atctgtaata atctagtcga ggcattcaag ccgacgtcgaccca ctctagaggat 3'	Complement of unique sequence k-complement of last 24 nucleotides of 3' UTR from RBCS2 gene	5' cgaatcctcgttag taactattccgactac caaatacgccca gcccgcccatgg 3'	SEQ ID NO 150
Nonshuffled segment V	Unique sequence k-First 25 nucleotides of the ble selectable marker cassette	5' gtagtcggaatagtt actaacgaggattcg gccagaaggag cgcagccaaaccag 3'	Complement of unique sequence l-complement of last 25 nucleotides of the ble selectable marker cassette	5' agttacgatttactag tcgagtagacattt ggtaccgggccc cccctcgagtta 3'	SEQ ID NO 149
Nonshuffled segment VI	Unique sequence l-first 24 nucleotides of promoter fragment of the lhcb1 gene	5' aaatgtctactcgac tagtaaatcgtaact gcagttgggtca ggggctggcgac 3'	Complement of unique sequence g-complement of last 25 nucleotides of promoter fragment of the lhcb1 gene	5' tcacacgattg ttaacgatttaag ccagtttaacgaaat gagtctcgcccgcggc 3'	SEQ ID NO 148
Nonshuffled segment VII	Unique sequence h-first 25 nucleotides of 3' UTR from RBCS2 gene	5' gatttaacat aactgtcgat taccgtgcga ccgacgtcgaccca ctctagaggat 3'	Complement of unique sequence i-complement of last 25 nucleotides of promoter fragment of the lhcb1 gene	5' ttgtcaccagga ttacgattgtcaagc atataacgaaatga gtctcgcccgcggc 3'	SEQ ID NO 151
Nonshuffled segment VIII	Unique sequence j-first 25 nucleotides of 3' UTR from RBCS2 gene	5' taacaagaat ctggctaatc aatcgatgca ccgacgtcgaccca ctctagaggat 3'	Complement of last 24 nucleotides of 3' UTR from RBCS2 gene	5' caaatacgccca gcccgcccatgg 3'	SEQ ID NO 150

[172] Table 2 Key to nomenclature: Chimeric oligonucleotides are designed according to sequences derived from the 5' and 3' ends of the 70 cDNAs of the 1-5 H_2 set. All portions of chimeric oligonucleotides corresponding to the 5' end of a cDNA start with a start codon. For instance, the oligonucleotide 1.1 from Table 1 has a sequence of 5' atccgtagttatccttatggccatcttagc-*atg[cpul1h2]*$_{27}$ 3'. This oligonucleotide's first 30 nucleotides, reading from 5' to 3', encode unique sequence a (SEQ ID NO 152). Nucleotides 31-33 encode a start codon (atg). After the start

codon the sequence is from the 5' end of the Chlamydomonas pulvinata 1 H_2 gene coding sequence, beginning after the start codon. Sequence listed in italics corresponds to the portion of the description written in italics. All portions of chimeric oligonucleotides corresponding to the 3' end of a cDNA end with a stop codon. For instance, the oligonucleotide 2.1 from Table 1 has a sequence of 5' [cpul1h2]$_{27}$taa-*cgtgcatcgattaacagcttctggacctga* 3'. This oligonucleotide's first 27 nucleotides, reading from 5' to 3', encode the last 27 nucleotides of the Chlamydomonas pulvinata 1 H_2 gene coding sequence, followed by a stop codon. After the stop codon the sequence is unique sequence b (SEQ ID NO 153).

TABLE 2

Oligo #	5' end corresponding to:	3' end corresponding to:	Sequence
1.1	Unique sequence a (SEQ ID NO 152)	*First 30 bp of 5' end of Chlamydomonas pulvinata 1 H_2 gene coding sequence*	5' atccgtagttatccttatggccatcttagc-*atg[cpul1h2]$_{27}$* 3'
1.2	Unique sequence a (SEQ ID NO 152)	*First 30 bp of 5' end of Chlamydomonas pygmaea 1 H_2 gene coding sequence*	5' atccgtagttatccttatggccatcttagc-*atg[cpyg1h2]$_{27}$* 3'
1.3	Unique sequence a (SEQ ID NO 152)	*First 30 bp of 5' end of Chlamydomonas radiata 1 H_2 gene coding sequence*	5' atccgtagttatccttatggccatcttagc-*atg[crad1h2]$_{27}$* 3'
1.4	Unique sequence a (SEQ ID NO 152)	*First 30 bp of 5' end of Chlamydomonas rapa 1 H_2 gene coding sequence*	5' atccgtagttatccttatggccatcttagc-*atg[crap1h2]$_{27}$* 3'
1.5	Unique sequence a (SEQ ID NO 152)	*First 30 bp of 5' end of Chlamydomonas sajao 1 H_2 gene coding sequence*	5' atccgtagttatccttatggccatcttagc-*atg[csaj1h2]$_{27}$* 3'
1.6	Unique sequence a (SEQ ID NO 152)	*First 30 bp of 5' end of Chlamydomonas segnis^{222} 1 H_2 gene coding sequence*	5' atccgtagttatccttatggccatcttagc-*atg[cseg^{222}1h2]$_{27}$* 3'
1.7	Unique sequence a (SEQ ID NO 152)	*First 30 bp of 5' end of Chlamydomonas segnis^{1638} 1 H_2 gene coding sequence*	5' atccgtagttatccttatggccatcttagc-*atg[cseg^{1638}1h2]$_{27}$* 3'
1.8	Unique sequence a (SEQ ID NO 152)	*First 30 bp of 5' end of Chlamydomonas segnis^{1919} 1 H_2 gene coding sequence*	5' atccgtagttatccttatggccatcttagc-*atg[cseg^{1919}1h2]$_{27}$* 3'
1.9	Unique sequence a (SEQ ID NO 152)	*First 30 bp of 5' end of Chlamydomonas smithii 1 H_2 gene coding sequence*	5' atccgtagttatccttatggccatcttagc-*atg[csmi1h2]$_{27}$* 3'
1.10	Unique sequence a (SEQ ID NO 152)	*First 30 bp of 5' end of Chlamydomonas sphaeroides 1 H_2 gene coding sequence*	5' atccgtagttatccttatggccatcttagc-*atg[csph1h2]$_{27}$* 3'
1.11	Unique sequence a (SEQ ID NO 152)	*First 30 bp of 5' end of Chlamydomonas surtseyiensis 1 H_2 gene coding sequence*	5' atccgtagttatccttatggccatcttagc-*atg[csur1h2]$_{27}$* 3'
1.12	Unique sequence a (SEQ ID NO 152)	*First 30 bp of 5' end of Chlamydomonas ulvaensis 1 H_2 gene coding sequence*	5' atccgtagttatccttatggccatcttagc-*atg[culv1h2]$_{27}$* 3'
1.13	Unique sequence a (SEQ ID NO 152)	*First 30 bp of 5' end of Chlamydomonas zimbabwiensis 1 H_2 gene coding sequence*	5' atccgtagttatccttatggccatcttagc-*atg[czim1h2]$_{27}$* 3'
1.14	Unique sequence a (SEQ ID NO 152)	*First 30 bp of 5' end of Chlamydomonas reinhardtii 1 H_2 gene coding sequence*	5' atccgtagttatccttatggccatcttagc-*atg[crei1h2]$_{27}$* 3'

2.1	Last 30 bp of 3' end of Chlamydomonas pulvinata 1 H_2 gene coding sequence	Unique sequence b (SEQ ID NO 153)	5' [cpul1h2]$_{30}$-*cgtgcatcga ttaacagcttctggacctga* 3'
2.2	Last 30 bp of 3' end of Chlamydomonas pygmaea 1 H_2 gene coding sequence	*Unique sequence b (SEQ ID NO 153)*	5' [cpyg1h2]$_{27}$taa-*cgtgcatcga ttaacagcttctggacctga* 3'
2.3	Last 30 bp of 3' end of Chlamydomonas radiata 1 H_2 gene coding sequence	*Unique sequence b (SEQ ID NO 153)*	5' [crad1h2] $_{27}$taa-*cgtgcatcga ttaacagcttctggacctga* 3'
2.4	Last 30 bp of 3' end of Chlamydomonas rapa 1 H_2 gene coding sequence	*Unique sequence b (SEQ ID NO 153)*	5' [crap1h2] $_{27}$taa-*cgtgcatcga ttaacagcttctggacctga* 3'
2.5	Last 30 bp of 3' end of Chlamydomonas sajao 1 H_2 gene coding sequence	*Unique sequence b (SEQ ID NO 153)*	5' [csaj1h2] $_{27}$taa-*cgtgcatcga ttaacagcttctggacctga* 3'
2.6	Last 30 bp of 3' end of Chlamydomonas segnis^{222} 1 H_2 gene coding sequence	*Unique sequence b (SEQ ID NO 153)*	5' [cseg^{222}1h2] $_{27}$taa-*cgtgcatcga ttaacagcttctggacctga* 3'
2.7	Last 30 bp of 3' end of Chlamydomonas segnis^{1638} 1 H_2 gene coding sequence	*Unique sequence b (SEQ ID NO 153)*	5' [cseg^{1638}1h2] $_{27}$taa-*cgtgcatcga ttaacagcttctggacctga* 3'
2.8	Last 30 bp of 3' end of Chlamydomonas segnis^{1919} 1 H_2 gene coding sequence	*Unique sequence b (SEQ ID NO 153)*	5' [cseg^{1919}1h2] $_{27}$taa-*cgtgcatcga ttaacagcttctggacctga* 3'
2.9	Last 30 bp of 3' end of Chlamydomonas smithii 1 H_2 gene coding sequence	*Unique sequence b (SEQ ID NO 153)*	5' [csmi1h2] $_{27}$taa-*cgtgcatcga ttaacagcttctggacctga* 3'
2.10	Last 30 bp of 3' end of Chlamydomonas sphaeroides 1 H_2 gene coding sequence	*Unique sequence b (SEQ ID NO 153)*	5' [csph1h2] $_{27}$taa -*cgtgcatcga ttaacagcttctggacctga* 3'
2.11	Last 30 bp of 3' end of Chlamydomonas surtseyiensis 1 H_2 gene coding sequence	*Unique sequence b (SEQ ID NO 153)*	5' [csur1h2] $_{27}$taa-*cgtgcatcga ttaacagcttctggacctga* 3'
2.12	Last 30 bp of 3' end of Chlamydomonas ulvaensis 1 H_2 gene coding sequence	*Unique sequence b (SEQ ID NO 153)*	5' [culv1h2] $_{27}$taa-*cgtgcatcga ttaacagcttctggacctga* 3'
2.13	Last 30 bp of 3' end of Chlamydomonas zimbabwiensis 1 H_2 gene coding sequence	*Unique sequence b (SEQ ID NO 153)*	5' [czim1h2] $_{27}$taa-*cgtgcatcga ttaacagcttctggacctga* 3'
2.14	Last 30 bp of 3' end of Chlamydomonas reinhardtii 1 H_2 gene coding sequence	*Unique sequence b (SEQ ID NO 153)*	5' [crei1h2] $_{27}$taa-*cgtgcatcga ttaacagcttctggacctga* 3'
3.1	Unique sequence c (SEQ ID NO 154)	*First 30 bp of 5' end of Chlamydomonas pulvinata 2 H_2 gene coding sequence*	5' ttaaacgtcgtacgtccaagtataactaag-*atg[cpul1h2]$_{27}$* 3'
3.2	Unique sequence c (SEQ ID NO 154)	*First 30 bp of 5' end of Chlamydomonas pygmaea 2 H_2 gene coding sequence*	5' ttaaacgtcgtacgtccaagtataactaag-*atg[cpyg1h2]$_{27}$* 3'

3.3	Unique sequence c (SEQ ID NO 154)	*First 30 bp of 5' end of Chlamydomonas radiata 2 H_2 gene coding sequence*	5' ttaaacgtcgtacgtccaagtataactaag-*atg[crad1h2]$_{27}$* 3'
3.4	Unique sequence c (SEQ ID NO 154)	*First 30 bp of 5' end of Chlamydomonas rapa 2 H_2 gene coding sequence*	5' ttaaacgtcgtacgtccaagtataactaag-*atg[crap1h2]$_{27}$* 3'
3.5	Unique sequence c (SEQ ID NO 154)	*First 30 bp of 5' end of Chlamydomonas sajao 2 H_2 gene coding sequence*	5' ttaaacgtcgtacgtccaagtataactaag-*atg[csaj1h2]$_{27}$* 3'
3.6	Unique sequence c (SEQ ID NO 154)	*First 30 bp of 5' end of Chlamydomonas segnis^{222} 2 H_2 gene coding sequence*	5' ttaaacgtcgtacgtccaagtataactaag-*atg[cseg^{222}1h2]$_{27}$* 3'
3.7	Unique sequence c (SEQ ID NO 154)	*First 30 bp of 5' end of Chlamydomonas segnis^{1638} 2 H_2 gene coding sequence*	5' ttaaacgtcgtacgtccaagtataactaag-*atg[cseg^{1638}1h2]$_{27}$* 3'
3.8	Unique sequence c (SEQ ID NO 154)	*First 30 bp of 5' end of Chlamydomonas segnis^{1919} 2 H_2 gene coding sequence*	5' ttaaacgtcgtacgtccaagtataactaag-*atg[cseg^{1919}1h2]$_{27}$* 3'
3.9	Unique sequence c (SEQ ID NO 154)	*First 30 bp of 5' end of Chlamydomonas smithii 2 H_2 gene coding sequence*	5' ttaaacgtcgtacgtccaagtataactaag-*atg[csmi1h2]$_{27}$* 3'
3.10	Unique sequence c (SEQ ID NO 154)	*First 30 bp of 5' end of Chlamydomonas sphaeroides 2 H_2 gene coding sequence*	5' ttaaacgtcgtacgtccaagtataactaag-*atg[csph1h2]$_{27}$* 3'
3.11	Unique sequence c (SEQ ID NO 154)	*First 30 bp of 5' end of Chlamydomonas surtseyiensis 2 H_2 gene coding sequence*	5' ttaaacgtcgtacgtccaagtataactaag-*atg[csur1h2]$_{27}$* 3'
3.12	Unique sequence c (SEQ ID NO 154)	*First 30 bp of 5' end of Chlamydomonas ulvaensis 2 H_2 gene coding sequence*	5' ttaaacgtcgtacgtccaagtataactaag-*atg[culv1h2]$_{27}$* 3'
3.13	Unique sequence c (SEQ ID NO 154)	*First 30 bp of 5' end of Chlamydomonas zimbabwiensis 2 H_2 gene coding sequence*	5' ttaaacgtcgtacgtccaagtataactaag-*atg[czim1h2]$_{27}$* 3'
3.14	Unique sequence c (SEQ ID NO 154)	*First 30 bp of 5' end of Chlamydomonas reinhardtii 2 H_2 gene coding sequence*	5' ttaaacgtcgtacgtccaagtataactaag-*atg[crei1h2]$_{27}$* 3'
4.1	Last 30 bp of 3' end of Chlamydomonas pulvinata 1 H_2 gene coding sequence	*Unique sequence d (SEQ ID NO 155)*	5' [cpul2h2]$_{27}$taa-*aatctgatac atgctattcagatcttacaa* 3'
4.2	Last 30 bp of 3' end of Chlamydomonas pygmaea 2 H_2 gene coding sequence	*Unique sequence d (SEQ ID NO 155)*	5' [cpyg2h2] $_{27}$taa-*aatctgatac atgctattcagatcttacaa* 3'
4.3	Last 30 bp of 3' end of Chlamydomonas radiata 2 H_2 gene coding sequence	*Unique sequence d (SEQ ID NO 155)*	5' [crad2h2] $_{27}$taa-*aatctgatac atgctattcagatcttacaa* 3'
4.4	Last 30 bp of 3' end of Chlamydomonas rapa 2 H_2 gene coding sequence	*Unique sequence d (SEQ ID NO 155)*	5' [crap2h2] $_{27}$taa-*aatctgatac atgctattcagatcttacaa* 3'
4.5	Last 30 bp of 3' end of Chlamydomonas sajao 2 H_2 gene coding sequence	*Unique sequence d (SEQ ID NO 155)*	5' [csaj2h2] $_{27}$taa-*aatctgatac atgctattcagatcttacaa* 3'
4.6	Last 30 bp of 3' end of Chlamydomonas segnis^{222} 2 H_2 gene coding sequence	*Unique sequence d (SEQ ID NO 155)*	5' [cseg^{222}2h2] $_{27}$taa-*aatctgatac atgctattcagatcttacaa* 3'

4.7	Last 30 bp of 3' end of Chlamydomonas segnis[1638] 2 H_2 gene coding sequence	*Unique sequence d (SEQ ID NO 155)*	5' [cseg[1638]2h2] $_{27}$taa-*aatctgatac atgctattcagatcttacaa* 3'
4.8	Last 30 bp of 3' end of Chlamydomonas segnis[1919] 2 H_2 gene coding sequence	*Unique sequence d (SEQ ID NO 155)*	5' [cseg[1919]2h2] $_{27}$taa-*aatctgatac atgctattcagatcttacaa* 3'
4.9	Last 30 bp of 3' end of Chlamydomonas smithii 2 H_2 gene coding sequence	*Unique sequence d (SEQ ID NO 155)*	5' [csmi2h2] $_{27}$taa-*aatctgatac atgctattcagatcttacaa* 3'
4.10	Last 30 bp of 3' end of Chlamydomonas sphaeroides 2 H_2 gene coding sequence	*Unique sequence d (SEQ ID NO 155)*	5' [csph2h2] $_{27}$taa-*aatctgatac atgctattcagatcttacaa* 3'
4.11	Last 30 bp of 3' end of Chlamydomonas surtseyiensis 2 H_2 gene coding sequence	*Unique sequence d (SEQ ID NO 155)*	5' [csur2h2] $_{27}$taa-*aatctgatac atgctattcagatcttacaa* 3'
4.12	Last 30 bp of 3' end of Chlamydomonas ulvaensis 2 H_2 gene coding sequence	*Unique sequence d (SEQ ID NO 155)*	5' [culv2h2] $_{27}$taa-*aatctgatac atgctattcagatcttacaa* 3'
4.13	Last 30 bp of 3' end of Chlamydomonas zimbabwiensis 2 H_2 gene coding sequence	*Unique sequence d (SEQ ID NO 155)*	5' [czim2h2] $_{27}$taa-*aatctgatac atgctattcagatcttacaa* 3'
4.14	Last 30 bp of 3' end of Chlamydomonas reinhardtii 2 H_2 gene coding sequence	*Unique sequence d (SEQ ID NO 155)*	5' [crei2h2] $_{27}$taa-*aatctgatac atgctattcagatcttacaa* 3'
5.1	Unique sequence e (SEQ ID NO 156)	*First 30 bp of 5' end of Chlamydomonas pulvinata 3 H_2 gene coding sequence*	5' aaatgtctactcgactagtaaatcgtaact-*atg[cpul3h2]$_{27}$* 3'
5.2	Unique sequence e (SEQ ID NO 156)	*First 30 bp of 5' end of Chlamydomonas pygmaea 3 H_2 gene coding sequence*	5' aaatgtctactcgactagtaaatcgtaact-*atg[cpyg3h2]$_{27}$* 3'
5.3	Unique sequence e (SEQ ID NO 156)	*First 30 bp of 5' end of Chlamydomonas radiata 3 H_2 gene coding sequence*	5' aaatgtctactcgactagtaaatcgtaact-*atg[crad3h2]$_{27}$* 3'
5.4	Unique sequence e (SEQ ID NO 156)	*First 30 bp of 5' end of Chlamydoimonas rapa 3 H_2 gene coding sequence*	5' aaatgtctactcgactagtaaatcgtaact-*atg[crap3h2]$_{27}$* 3'
5.5	Unique sequence e (SEQ ID NO 156)	*First 30 bp of 5' end of Chlamydomonas sajao 3 H_2 gene coding sequence*	5' aaatgtctactcgactagtaaatcgtaact-*atg[csaj3h2]$_{27}$* 3'
5.6	Unique sequence e (SEQ ID NO 156)	*First 30 bp of 5' end of Chlamydomonas segnis[222] 3 H_2 gene coding sequence*	5' aaatgtctactcgactagtaaatcgtaact-*atg[cseg[222]3h2]$_{27}$* 3'
5.7	Unique sequence e (SEQ ID NO 156)	*First 30 bp of 5' end of Chlamydomonas segnis[1638] 3 H_2 gene coding sequence*	5' aaatgtctactcgactagtaaatcgtaact-*atg[cseg[1638]3h2]$_{27}$* 3'
5.8	Unique sequence e (SEQ ID NO 156)	*First 30 bp of 5' end of Chlamydomonas segnis[1919] 3 H_2 gene coding sequence*	5' aaatgtctactcgactagtaaatcgtaact-*atg[cseg[1919]3h2]$_{27}$* 3'
5.9	Unique sequence e (SEQ ID NO 156)	*First 30 bp of 5' end of Chlamydomonas smithii 3 H_2 gene coding sequence*	5' aaatgtctactcgactagtaaatcgtaact-*atg[csmi3h2]$_{27}$* 3'

5.10	Unique sequence e (SEQ ID NO 156)	*First 30 bp of 5' end of Chlamydomonas sphaeroides 3 H_2 gene coding sequence*	5' aaatgtctactcgactagtaaatcgtaact-*atg[csph3h2]$_{27}$* 3'
5.11	Unique sequence e (SEQ ID NO 156)	*First 30 bp of 5' end of Chlamydomonas surtseyiensis 3 H_2 gene coding sequence*	5' aaatgtctactcgactagtaaatcgtaact-*atg[csur3h2]$_{27}$* 3'
5.12	Unique sequence e (SEQ ID NO 156)	*First 30 bp of 5' end of Chlamydomonas ulvaensis 3 H_2 gene coding sequence*	5' aaatgtctactcgactagtaaatcgtaact-*atg[culv3h2]$_{27}$* 3'
5.13	Unique sequence e (SEQ ID NO 156)	*First 30 bp of 5' end of Chlamydomonas zimbabwiensis 3 H_2 gene coding sequence*	5' aaatgtctactcgactagtaaatcgtaact-*atg[czim3h2]$_{27}$* 3'
5.14	Unique sequence e (SEQ ID NO 156)	*First 30 bp of 5' end of Chlamydomonas reinhardtii 3 H_2 gene coding sequence*	5' aaatgtctactcgactagtaaatcgtaact-*atg[crei3h2]$_{27}$* 3'
6.1	Last 30 bp of 5' end of Chlamydomonas pulvinata 3 H_2 gene coding sequence	*Unique sequence f (SEQ ID NO 157)*	5' [cpul3h2]$_{27}$taa-*atctgtaata atctagtcgaggcattcaag* 3'
6.2	Last 30 bp of 3' end of Chlamydomonas pygmaea 3 H_2 gene coding sequence	*Unique sequence f (SEQ ID NO 157*	5' [cpyg3h2] $_{27}$taa-*atctgtaata atctagtcgaggcattcaag* 3'
6.3	Last 30 bp of 3' end of Chlamydomonas radiata 3 H_2 gene coding sequence	*Unique sequence f (SEQ ID NO 157*	5' [crad3h2] $_{27}$taa-*atctgtaata atctagtcgaggcattcaag* 3'
6.4	Last 30 bp of 3' end of Chlamydomonas rapa 3 H_2 gene coding sequence	*Unique sequence f (SEQ ID NO 157*	5' [crap3h2] $_{27}$taa-*atctgtaata atctagtcgaggcattcaag* 3'
6.5	Last 30 bp of 3' end of Chlamydomonas sajao 3 H_2 gene coding sequence	*Unique sequence f (SEQ ID NO 157*	5' [csaj3h2] $_{27}$taa-*atctgtaata atctagtcgaggcattcaag* 3'
6.6	Last 30 bp of 3' end of Chlamydomonas segnis^{222} 3 H_2 gene coding sequence	*Unique sequence f (SEQ ID NO 157*	5' [cseg^{222}3h2] $_{27}$taa-*atctgtaata atctagtcgaggcattcaag* 3'
6.7	Last 30 bp of 3' end of Chlamydomonas segnis^{1638} 3 H_2 gene coding sequence	*Unique sequence f (SEQ ID NO 157*	5' [cseg^{1638}3h2] $_{27}$taa-*atctgtaata atctagtcgaggcattcaag* 3'
6.8	Last 30 bp of 3' end of Chlamydomonas segnis^{1919} 3 H_2 gene coding sequence	*Unique sequence f (SEQ ID NO 157*	5' [cseg^{1919}3h2] $_{27}$taa-*atctgtaata atctagtcgaggcattcaag* 3'
6.9	Last 30 bp of 3' end of Chlamydomonas smithii 3 H_2 gene coding sequence	*Unique sequence f (SEQ ID NO 157*	5' [csmi3h2] $_{27}$taa-*atctgtaata atctagtcgaggcattcaag* 3'
6.10	Last 30 bp of 3' end of Chlamydomonas sphaeroides 3 H_2 gene coding sequence	*Unique sequence f (SEQ ID NO 157*	5' [csph3h2] $_{27}$taa-*atctgtaata atctagtcgaggcattcaag* 3'
6.11	Last 30 bp of 3' end of Chlamydomonas surtseyiensis 3 H_2 gene coding sequence	*Unique sequence f (SEQ ID NO 157*	5' [csur3h2] $_{27}$taa-*atctgtaata atctagtcgaggcattcaag* 3'

6.12	Last 30 bp of 3' end of Chlamydomonas ulvaensis 3 H_2 gene coding sequence	*Unique sequence f (SEQ ID NO 157*	5' [culv3h2] $_{27}$taa-*atctgtaata atctagtcgaggcattcaag* 3'
6.13	Last 30 bp of 3' end of Chlamydomonas zimbabwiensis 3 H_2 gene coding sequence	*Unique sequence f (SEQ ID NO 157*	5' [czim3h2] $_{27}$taa-*atctgtaata atctagtcgaggcattcaag* 3'
6.14	Last 30 bp of 3' end of Chlamydomonas reinhardtii 3 H_2 gene coding sequence	*Unique sequence f (SEQ ID NO 157*	5' [crei3h2] $_{27}$taa-*atctgtaata atctagtcgaggcattcaag* 3'
7.1	Unique sequence g (SEQ ID NO 158)	*First 30 bp of 5' end of Chlamydomonas pulvinata 4 H_2 gene coding sequence*	5' aactggcttaaatcgttaacaatcgtgtga-*atg[cpul4h2]$_{27}$* 3'
7.2	Unique sequence g (SEQ ID NO 158)	*First 30 bp of 5' end of Chlamydomonas pygmaea 4 H_2 gene coding sequence*	5' aactggcttaaatcgttaacaatcgtgtga-*atg[cpyg4h2]$_{27}$* 3'
7.3	Unique sequence g (SEQ ID NO 158)	*First 30 bp of 5' end of Chlamydomonas radiata 4 H_2 gene coding sequence*	5' aactggcttaaatcgttaacaatcgtgtga-*atg[crad4h2]$_{27}$* 3'
7.4	Unique sequence g (SEQ ID NO 158)	*First 30 bp of 5' end of Chlamydomonas rapa 4 H_2 gene coding sequence*	5' aactggcttaaatcgttaacaatcgtgtga-*atg[crap4h2]$_{27}$* 3'
7.5	Unique sequence g (SEQ ID NO 158)	*First 30 bp of 5' end of Chlamydomonas sajao 4 H_2 gene coding sequence*	5' aactggcttaaatcgttaacaatcgtgtga-*atg[csaj4h2]$_{27}$* 3'
7.6	Unique sequence g (SEQ ID NO 158)	*First 30 bp of 5' end of Chlamydomonas segnis^{222} 4 H_2 gene coding sequence*	5' aactggcttaaatcgttaacaatcgtgtga-*atg[cseg^{222}4h2]$_{27}$* 3'
7.7	Unique sequence g (SEQ ID NO 158)	*First 30 bp of 5' end of Chlamydomonas segnis^{1638} 4 H_2 gene coding sequence*	5' aactggcttaaatcgttaacaatcgtgtga-*atg[cseg^{1638}4h2]$_{27}$* 3'
7.8	Unique sequence g (SEQ ID NO 158)	*First 30 bp of 5' end of Chlamydomonas segnis^{1919} 4 H_2 gene coding sequence*	5' aactggcttaaatcgttaacaatcgtgtga-*atg[cseg^{1919}4h2]$_{27}$* 3'
7.9	Unique sequence g (SEQ ID NO 158)	*First 30 bp of 5' end of Chlamydomonas smithii 4 H_2 gene coding sequence*	5' aactggcttaaatcgttaacaatcgtgtga-*atg[csmi4h2]$_{27}$* 3'
7.10	Unique sequence g (SEQ ID NO 158)	*First 30 bp of 5' end of Chlamydomonas sphaeroides 4 H_2 gene coding sequence*	5' aactggcttaaatcgttaacaatcgtgtga-*atg[csph4h2]$_{27}$* 3'
7.11	Unique sequence g (SEQ ID NO 158)	*First 30 bp of 5' end of Chlamydomonas surtseyiensis 4 H_2 gene coding sequence*	5' aactggcttaaatcgttaacaatcgtgtga-*atg[csur4h2]$_{27}$* 3'
7.12	Unique sequence g (SEQ ID NO 158)	*First 30 bp of 5' end of Chlamydomonas ulvaensis 4 H_2 gene coding sequence*	5' aactggcttaaatcgttaacaatcgtgtga-*atg[culv4h2]$_{27}$* 3'
7.13	Unique sequence g (SEQ ID NO 158)	*First 30 bp of 5' end of Chlamydomonas zimbabwiensis 4 H_2 gene coding sequence*	5' aactggcttaaatcgttaacaatcgtgtga-*atg[czim4h2]$_{27}$* 3'
7.14	Unique sequence g (SEQ ID NO 158)	*First 30 bp of 5' end of Chlamydomonas reinhardtii 4 H_2 gene coding sequence*	5' aactggcttaaatcgttaacaatcgtgtga-*atg[crei4h2]$_{27}$* 3'
8.1	Last 30 bp of 5' end of Chlamydomonas pulvinata 4 H_2 gene coding sequence	*Unique sequence h (SEQ ID NO 159)*	5' [cpul4h2]$_{27}$taa-*gatttaacat aactgtcgattaccgtgcga* 3'

8.2	Last 30 bp of 3' end of Chlamydomonas pygmaea 4 H_2 gene coding sequence	*Unique sequence h (SEQ ID NO 159)*	5' [cpyg4h2] $_{27}$taa-*gatttaacat aactgtcgattaccgtgcga* 3'
8.3	Last 30 bp of 3' end of Chlamydomonas radiata 4 H_2 gene coding sequence	*Unique sequence h (SEQ ID NO 159)*	5' [crad4h2] $_{27}$taa-*gatttaacat aactgtcgattaccgtgcga* 3'
8.4	Last 30 bp of 3' end of Chlamydomonas rapa 4 H_2 gene coding sequence	*Unique sequence h (SEQ ID NO 159)*	5' [crap4h2] $_{27}$taa-*gatttaacat aactgtcgattaccgtgcga* 3'
8.5	Last 30 bp of 3' end of Chlamydomonas sajao 4 H_2 gene coding sequence	*Unique sequence h (SEQ ID NO 159)*	5' [csaj4h2] $_{27}$taa-*gatttaacat aactgtcgattaccgtgcga* 3'
8.6	Last 30 bp of 3' end of Chlamydomonas segnis^{222} 4 H_2 gene coding sequence	*Unique sequence h (SEQ ID NO 159)*	5' [cseg^{222}4h2] $_{27}$taa-*gatttaacat aactgtcgattaccgtgcga* 3'
8.7	Last 30 bp of 3' end of Chlamydomonas segnis^{1638} 4 H_2 gene coding sequence	*Unique sequence h (SEQ ID NO 159)*	5' [cseg^{1638}4h2] $_{27}$taa-*gatttaacat aactgtcgattaccgtgcga* 3'
8.8	Last 30 bp of 3' end of Chlamydomonas segnis^{1919} 4 H_2 gene coding sequence	*Unique sequence h (SEQ ID NO 159)*	5' [cseg^{1919}4h2] $_{27}$taa-*gatttaacat aactgtcgattaccgtgcga* 3'
8.9	Last 30 bp of 3' end of Chlamydomonas smithii 4 H_2 gene coding sequence	*Unique sequence h (SEQ ID NO 159)*	5' [csmi4h2] $_{27}$taa-*gatttaacat aactgtcgattaccgtgcga* 3'
8.10	Last 30 bp of 3' end of Chlamydomonas sphaeroides 4 H_2 gene coding sequence	*Unique sequence h (SEQ ID NO 159)*	5' [csph4h2] $_{27}$taa-*gatttaacat aactgtcgattaccgtgcga* 3'
8.11	Last 30 bp of 3' end of Chlamydomonas surtseyiensis 4 H_2 gene coding sequence	*Unique sequence h (SEQ ID NO 159)*	5' [csur4h2] $_{27}$taa-*gatttaacat aactgtcgattaccgtgcga* 3'
8.12	Last 30 bp of 3' end of Chlamydomonas ulvaensis 4 H_2 gene coding sequence	*Unique sequence h (SEQ ID NO 159)*	5' [culv4h2] $_{27}$taa-*gatttaacat aactgtcgattaccgtgcga* 3'
8.13	Last 30 bp of 3' end of Chlamydomonas zimbabwiensis 4 H_2 gene coding sequence	*Unique sequence h (SEQ ID NO 159)*	5' [czim4h2] $_{27}$taa-*gatttaacat aactgtcgattaccgtgcga* 3'
8.14	Last 30 bp of 3' end of Chlamydomonas reinhardtii 4 H_2 gene coding sequence	*Unique sequence h (SEQ ID NO 159)*	5' [crei4h2] $_{27}$taa-*gatttaacat aactgtcgattaccgtgcga* 3'
9.1	Unique sequence i (SEQ ID NO 160)	*First 30 bp of 5' end of Chlamydomonas pulvinata 5 H_2 gene coding sequence*	5' tatgcttgacaatcgtaatcctggtgacaa-*atg[cpul5h2]$_{27}$* 3'
9.2	Unique sequence i (SEQ ID NO 160)	*First 30 bp of 5' end of Chlamydomonas pygmaea 5 H_2 gene coding sequence*	5' tatgcttgacaatcgtaatcctggtgacaa-*atg[cpyg5h2]$_{27}$* 3'
9.3	Unique sequence i (SEQ ID NO 160)	*First 30 bp of 5' end of Chlamydomonas radiata 5 H_2 gene coding sequence*	5' tatgcttgacaatcgtaatcctggtgacaa-*atg[crad5h2]$_{27}$* 3'

9.4	Unique sequence i (SEQ ID NO 160)	*First 30 bp of 5' end of Chlamydomonas rapa 5 H_2 gene coding sequence*	5' tatgcttgacaatcgtaatcctggtgacaa-*atg[crap5h2]$_{27}$* 3'
9.5	Unique sequence i (SEQ ID NO 160)	*First 30 bp of 5' end of Chlamydomonas sajao 5 H_2 gene coding sequence*	5' tatgcttgacaatcgtaatcctggtgacaa-*atg[csaj5h2]$_{27}$* 3'
9.6	Unique sequence i (SEQ ID NO 160)	*First 30 bp of 5' end of Chlamydomonas segnis^{222} 5 H_2 gene coding sequence*	5' tatgcttgacaatcgtaatcctggtgacaa-*atg[cseg^{222}5h2]$_{27}$* 3'
9.7	Unique sequence i (SEQ ID NO 160)	*First 30 bp of 5' end of Chlamydomonas segnis^{1638} 5 H_2 gene coding sequence*	5' tatgcttgacaatcgtaatcctggtgacaa-*atg[cseg^{1638}5h2]$_{27}$* 3'
9.8	Unique sequence i (SEQ ID NO 160)	*First 30 bp of 5' end of Chlamydomonas segnis^{1919} 5 H_2 gene coding sequence*	5' tatgcttgacaatcgtaatcctggtgacaa-*atg[cseg^{1919}5h2]$_{27}$* 3'
9.9	Unique sequence i (SEQ ID NO 160)	*First 30 bp of 5' end of Chlamydomonas smithii 5 H_2 gene coding sequence*	5' tatgcttgacaatcgtaatcctggtgacaa-*atg[csmi5h2]$_{27}$* 3'
9.10	Unique sequence i (SEQ ID NO 160)	*First 30 bp of 5' end of Chlamydomonas sphaeroides 5 H_2 gene coding sequence*	5' tatgcttgacaatcgtaatcctggtgacaa-*atg[csph5h2]$_{27}$* 3'
9.11	Unique sequence i (SEQ ID NO 160)	*First 30 bp of 5' end of Chlamydomonas surtseyiensis 5 H_2 gene coding sequence*	5' tatgcttgacaatcgtaatcctggtgacaa-*atg[csur5h2]$_{27}$* 3'
9.12	Unique sequence i (SEQ ID NO 160)	*First 30 bp of 5' end of Chlamydomonas ulvaensis 5 H_2 gene coding sequence*	5' tatgcttgacaatcgtaatcctggtgacaa-*atg[culv5h2]$_{27}$* 3'
9.13	Unique sequence i (SEQ ID NO 160)	*First 30 bp of 5' end of Chlamydomonas zimbabwiensis 5 H_2 gene coding sequence*	5' tatgcttgacaatcgtaatcctggtgacaa-*atg[czim5h2]$_{27}$* 3'
9.14	Unique sequence i (SEQ ID NO 160)	*First 30 bp of 5' end of Chlamydomonas reinhardtii 5 H_2 gene coding sequence*	5' tatgcttgacaatcgtaatcctggtgacaa-*atg[crei5h2]$_{27}$* 3'
10.1	Last 30 bp of 5' end of Chlamydomonas pulvinata 5 H_2 gene coding sequence	*Unique sequence j (SEQ ID NO 161)*	5' [cpul5h2]$_{30}$-*taacaagaat ctggctaatcaatcgatgca* 3'
10.2	Last 30 bp of 3' end of Chlamydomonas pygmaea 5 H_2 gene coding sequence	*Unique sequence j (SEQ ID NO 161)*	5' [cpyg5h2]$_{27}$taa-*taacaagaat ctggctaatcaatcgatgca* 3'
10.3	Last 30 bp of 3' end of Chlamydomonas radiata 5 H_2 gene coding sequence	*Unique sequence j (SEQ ID NO 161)*	5' [crad5h2] $_{27}$taa-*taacaagaat ctggctaatcaatcgatgca* 3'
10.4	Last 30 bp of 3' end of Chlamydomonas rapa 5 H_2 gene coding sequence	*Unique sequence j (SEQ ID NO 161)*	5' [crap5h2] $_{27}$taa-*taacaagaat ctggctaatcaatcgatgca* 3'
10.5	Last 30 bp of 3' end of Chlamydomonas sajao 5 H_2 gene coding sequence	*Unique sequence j (SEQ ID NO 161)*	5' [csaj5h2] $_{27}$taa-*taacaagaat ctggctaatcaatcgatgca* 3'
10.6	Last 30 bp of 3' end of Chlamydomonas segnis^{222} 5 H_2 gene coding sequence	*Unique sequence j (SEQ ID NO 161)*	5' [cseg^{222}5h2] $_{27}$taa-*taacaagaat ctggctaatcaatcgatgca* 3'
10.7	Last 30 bp of 3' end of Chlamydomonas segnis^{1638} 5 H_2 gene coding sequence	*Unique sequence j (SEQ ID NO 161)*	5' [cseg^{1638}5h2] $_{27}$taa-*taacaagaat ctggctaatcaatcgatgca* 3'

10.8	Last 30 bp of 3' end of Chlamydomonas segnis[1919] 5 H_2 gene coding sequence	*Unique sequence j (SEQ ID NO 161)*	5' [cseg[1919]5h2] $_{27}$taa-*taacaagaat ctggctaatcaatcgatgca* 3'
10.9	Last 30 bp of 3' end of Chlamydomonas smithii 5 H_2 gene coding sequence	*Unique sequence j (SEQ ID NO 161)*	5' [csmi5h2] $_{27}$taa-*taacaagaat ctggctaatcaatcgatgca* 3'
10.10	Last 30 bp of 3' end of Chlamydomonas sphaeroides 5 H_2 gene coding sequence	*Unique sequence j (SEQ ID NO 161)*	5' [csph5h2] $_{27}$taa-*taacaagaat ctggctaatcaatcgatgca* 3'
10.11	Last 30 bp of 3' end of Chlamydomonas surtseyiensis 5 H_2 gene coding sequence	*Unique sequence j (SEQ ID NO 161)*	5' [csur5h2] $_{27}$taa-*taacaagaat ctggctaatcaatcgatgca* 3'
10.12	Last 30 bp of 3' end of Chlamydomonas ulvaensis 5 H_2 gene coding sequence	*Unique sequence j (SEQ ID NO 161)*	5' [culv5h2] $_{27}$taa-*taacaagaat ctggctaatcaatcgatgca* 3'
10.13	Last 30 bp of 3' end of Chlamydomonas zimbabwiensis 5 H_2 gene coding sequence	*Unique sequence j (SEQ ID NO 161)*	5' [czim5h2] $_{27}$taa-*taacaagaat ctggctaatcaatcgatgca* 3'
10.14	Last 30 bp of 3' end of Chlamydomonas reinhardtii 5 H_2 gene coding sequence	*Unique sequence j (SEQ ID NO 161)*	5' [crei5h2] $_{27}$taa-*taacaagaat ctggctaatcaatcgatgca* 3'

TABLE 3

Product	5' primer	5' primer sequence	3' primer	3' primer sequence	Template
Nonshuffled segment IX	Unique sequence a-First 24 nucleotides of promoter fragment of the lhcb1 gene	5' atccgtagtt atccttatgg ccatcttagc gcagttgggtca ggggctggcgac 3'	Complement of unique sequence b-complement of last 25 base pairs of the promoter fragment of the lhcb1 gene	5' tcaggtccagaagctgt taatcgatgcacg taacgaaatgag tctcgcccgcggc3'	SEQ ID NO 148
Nonshuffled segment X	Unique sequence c-first 25 nucleotides of 3' UTR from RBCS2 gene	5' ttaaacgtcg tacgtccaag tataactaag ccgacgtcgaccca ctctagaggat 3'	Complement of unique sequence d-complement of last 25 base pairs of the promoter fragment of the lhcb1 gene	5' ttgtaagatctgaat agcatgtatcagatt taacgaaatgag tctcgcccgcggc 3'	SEQ ID NO 151

Nonshuffled segment XI	Unique sequence e-first 25 nucleotides of 3' UTR from RBCS2 gene	5' tcttccatcg taaatctagc atcgattagc ccgacgtcgaccca ctctagaggat 3'	Complement of unique sequence f-complement of last 25 base pairs of the promoter fragment of the lhcb1 gene	5' cttgaatgcctcgact agattattacagat taacgaaatgag tctcgcccgcggc 3'	SEQ ID NO 151
Nonshuffled segment XII	Unique sequence f-first 25 nucleotides of synthetic green fluorescent protein gene (SEQ ID NO 32)	5' atctgtaataatctag tcgaggcattcaag atggccaagggcga ggagctgttca 3'	Complement of unique sequence g-complement of last 25 nucleotides of synthetic green fluorescent protein gene	5' tcacacgattgttaa cgatttaagccagtt ttacttgtacagctcg tccatgccg 3'	SEQ ID NO 179
Nonshuffled segment XIII	Unique sequence g-first 25 nucleotides of 3' UTR from RBCS2 gene	5' aactggctta aatcgttaac aatcgtgtga ccgacgtcgaccca ctctagaggat 3'	Complement of unique sequence h-Complement of last 24 nucleotides of 3' UTR from RBCS2 gene	5' tcgcacggtaatcgac agttatgttaaatc caaatacgcccagcc cgcccatgga 3'	SEQ ID NO 150

TABLE 4

Oligo #	5' end corresponding to:	3' end corresponding to:	Sequence
11.1	Unique sequence b	*First 25 nucleotides of Chlamydomonas reinhardtii hydrogenase*	5' cgtgcatcgattaacagcttctggacctga *atgtcggcgctcgtgctgaagccct* 3'
11.2	Unique sequence b	*First 25 nucleotides of Clostriduim pasteuranum hydrogenase*	5' cgtgcatcgattaacagcttctggacctga *atgaaaacaataattataaatggtg* 3'
11.3	Unique sequence b	*First 25 nucleotides of Desulfovibrio vulgaris hydrogenase*	5' cgtgcatcgattaacagcttctggacctga *atgagccgtaccgtcatggagcgca* 3'
11.4	Unique sequence b	*First 25 nucleotides of Entamoeba histolytica hydrogenase*	5' cgtgcatcgattaacagcttctggacctga *atgccacctaaaccatcacatacac* 3'
11.5	Unique sequence b	*First 25 nucleotides of Scenedesmus obliquus hydrogenase*	5' cgtgcatcgattaacagcttctggacctga *atgcctgagtggcaaccgggaggtc* 3'
11.6	Unique sequence b	*First 25 nucleotides of Chlorella fusca hydrogenase*	5' cgtgcatcgattaacagcttctggacctga *atgtgttgccccgtggttgcaagta* 3'
12.1	Complement of unique sequence c	*Complement of last 25 nucleotides of Chlamydomonas reinhardtii hydrogenase*	5' cttagttatacttggacgtacgacgtttaa *tcacttcttctcgtccttctcctcc* 3'
12.2	Complement of unique sequence c	*Complement of last 25 nucleotides of Clostriduim pasteuranum hydrogenase*	5' cttagttatacttggacgtacgacgtttaa *ttatttttatatttaaagtgtaat* 3'
12.3	Complement of unique sequence c	*Complement of last 25 nucleotides of Desulfovibrio vulgaris hydrogenase*	5' cttagttatacttggacgtacgacgtttaa *ctatgccttgttggcgctcgccatg* 3'
12.4	Complement of unique sequence c	*Complement of last 25 nucleotides of Entamoeba histolytica hydrogenase*	5' cttagttatacttggacgtacgacgtttaa *ttagttttgatatctgggagtaaaa* 3'

12.5	Complement of unique sequence c	*Complement of last 25 nucleotides of Scenedesmus obliquus hydrogenase*	5' cttagttatacttggacgtacgacgtttaa *tcacttctcatcgggcacgccgccg* 3'
12.6	Complement of unique sequence c	*Complement of last 25 nucleotides of Chlorella fusca hydrogenase*	5' cttagttatacttggacgtacgacgtttaa *tcacttctcctctggaattccacct* 3'
14	Unique sequence d	*First 25 nucleotides of Chlamydomonas reinhardtii ferredoxin*	5' aatctgatacatgctattcagatcttacaa *atggccatggctatgcgctccacct* 3'
15	Complement of unique sequence e	*Complement of last 25 nucleotides of Chlamydomonas reinhardtii ferredoxin*	5' gctaatcgatgctagatttacgatggaaga *ttagtacagggcctcctcctggtgg* 3'

U.S. PATENTS REFERENCED

Other patents included in paragraph **[073]** are 5,537,776; 5,965,408; 6,171,820; 6,174,673; 6,238,884; 6,326,204; 6,344,328; 6,352,842; 6,358,709; 6361,97; 6,368,798; 6,440,668; 6537,776; and 6,605,449.

Other patents referenced in this application are U.S. Patents 5,871,952, 5,605,79, 5,830,721, 6,165,793, 6,180,406, 5,939,250, 6,171,820, 6,361,974, 6,358,709, 6,352,842, 6,238,884 6,420,175, 6,287,861 , 6,277,589 , 4,532,210 and WO 01/48185 (Fischer).

WHAT IS CLAIMED IS:

1. A method for engineering a cell to produce an increased amount of hydrogen comprising:
   (a) providing a mutagenized nucleic acid sequence derived from a first gene that encodes a protein involved in a hydrogen production pathway;
   (b) transforming a cell with said mutagenized nucleic acid sequence; and
   (c) screening or selecting the cell for an increased amount of hydrogen.

2. The method of claim 1, wherein a plurality of mutagenized nucleic acid sequences are used to transform a population of cells, followed by the screening or selecting.

3. The method of claim 1, wherein the first gene is selected from the group that encodes ferredoxin, catalase, isoamylase, malate dehydrogenase, 14-3-3 protein, enolase, aldolase, ribosomal protein S8, ribosomal protein L17, ribosomal protein S18, ribosomal protein L37, ribosomal protein L12, ribosomal protein S15, iron-hydrogenase, nickel-iron hydrogenase, and components of the photosystem I, photosystem II, light harvesting antenna and cytochrome b_6-f complexes.

4. The method of claim 3, wherein the first gene encodes an iron-hydrogenase.

5. The method of claim 4, wherein at least one amino acid from the segment $X^1X^2X^3FX^4X^5X^6GGVMEAAX^7R$ or the segment ADX^8TIX^9EE is substituted by a different amino acid in the protein encoded by the first gene to generate the mutagenized nucleic acid sequence.

6. The method of claim 5, wherein the mutagenized nucleic acid sequence is generated by gene reassembly.

7. The method of claim 5, wherein the mutagenized nucleic acid sequence is generated by site-directed mutagenesis.

8. The method of claim 5, wherein an amino acid that is substituted for the at least one amino acid has a side chain of higher molecular weight than the side chain of the at least one amino acid.

9. The method of claim 5, wherein saturation mutagenesis is performed on the at least one amino acid.

10. The method of claim 5, wherein the mutagenized nucleic acid sequence is generated by a mutagenesis method described in U.S. Patents selected from the group consisting of 5,537,776; 5,965,408; 6,171,820; 6,174,673; 6,238,884; 6,326,204; 6, 344,328; 6,352,842; 6,358,709; 6361,97; 6,368,798; 6,440,668; 6537,776; and 6,605,449.

11. The method of claim 6, wherein the gene reassembly is performed using nucleic acid molecules that encode proteins of SEQ ID NOs: 1-112 or segments thereof.

12. The method of claim 4, wherein the mutagenized nucleic acid sequence encodes an iron hydrogenase protein that functionally interacts with a ferredoxin protein in the cell.

13. The method of claim 1, wherein the screening or selecting occurs in the presence of oxygen at a concentration selected from the ranges comprising more than 0.5%, more than 5.0%, more than 10%, more than 15%, approximately 21%, more than 21%, more than 25%, more than 30% or more than 35% oxygen.

14. The method of claim 1, wherein the mutagenized nucleic acid sequence is operably linked to a promoter that is activated by light.

15. The method of claim 1, wherein the mutagenized nucleic acid sequence is generated by gene reassembly.

16. The method of claim 1, wherein the cell is a green algae species.

17. The method of claim 1, wherein cell is of the genus Chlamydomonas.

18. The method of claim 1, further comprising the steps of;

(a) identifying a first independent transformant which produces an increased amount of hydrogen from step (c) of claim 1;

(b) recovering the mutagenized nucleic acid sequence from the independent transformant;

(c) further mutagenizing the recovered mutagenized nucleic acid sequence to create a new library of mutagenized nucleic acid sequences;

(d) transforming cells with the new library of mutagenized nucleic acid sequences; and

(e) screening or selecting for a new independent transformant from the new library that generates an increased amount of hydrogen compared to the first independent transformant.

19. The method of claim 18 wherein the mutagenized nucleic acid sequencs are generated by gene reassembly.

20. The method of claim 18, wherein a plurality of mutagenized nucleic acid sequences are recovered from a plurality of independent transformants which produce an increased amount of hydrogen from step (c) of claim 1, and wherein the plurality of mutagenized nucleic acid sequences are subjected to gene reassembly to generate the new library.

21. The method of claim 1, wherein the screening or selecting occurs by culturing cells in liquid growth media.

22. The method of claim 21, wherein the growth media is a photoautotrophic growth-requiring minimal media.

23. The method of claim 1, wherein the screening or selecting occurs in a non-transparent culture container.

24. A method according to claim 1, wherein the mutagenized nucleic acid sequence is operably linked to a promoter that is constitutively activated.

25. The method of claim 15, wherein the mutagenized nucleic acid sequence is obtained by

subjecting nucleic acid sequences that encode proteins that are expressed when cells are exposed to conditions more conducive to the generation of hydrogen to gene reassembly, wherein the proteins are naturally encoded by genes in organisms from more than one species.

26. The method of claim 19, wherein the proteins are iron hydrogenases or nickel-iron hydrogenases.

27. The method of claim 1, further comprising repeating the steps of claim 1 using a second gene distinct from the first gene.

28. The method of claim 27, further comprising:

(a) mating at least one cell of a strain containing a mutagenized form of the first gene:

i. wherein the at least one cell is identified by the screening or selecting; or

ii. wherein the at least one cell is derived through mating from a cell identified by the screening or selecting;

to at least one cell of a distinct strain containing a mutagenized form of the second gene:

iii. wherein the at least one cell is identified by the screening or selecting; or

iv. wherein the at least one cell is derived through mating from a cell identified by the screening or selecting; and

(b) screening or selecting for a progeny cell that produces an increased amount of hydrogen compared to any parent cell.

29. A method of hydrogen production, comprising:

(a) placing cell containing a mutagenized nucleic acid sequence corresponding to a gene that is involved in a hydrogen production pathway into liquid culture media or on to solid culture media, wherein the mutagenized nucleic acid sequence is operably linked to a transcriptional promoter sequence;

(b) culturing said transformed cell under conditions sufficient to stimulate transcription of said mutagenized nucleic acid sequence(s); and

(c) collecting an evolved gas.

30. The method of claim 29, wherein the culture media is photoautotrophic growth requiring media.

31. A method of multiparental mating of microbes that mate in response to a stimulus, comprising:

(a) providing a cell from each of 3 or more strains of microbes capable of mating to each other in culture medium;

(b) providing the stimulus;

(c) allowing cells to mate and produce progeny;

(d) allowing the progeny cells to achieve sexual reproduction capability;

(e) providing the stimulus at least one more time; and

(f) screening or selecting the further progeny for a desired phenotype.

32. The method of claim 31, wherein the microbes are green algae and the stimulus is the removal of

nitrogen from the media and illumination by light comprising a wavelength between about 0.42-0.52 micrometers.

33. The method of claim 32, wherein the green algae are of the Chlamydomonas genus.

34. The method of claim 33, wherein the species is selected from the group comprising reinhardtii, eugametos, incerta, and moewusii.

35. The method of claim 31, wherein the stimulus is interruption of exponential growth in continuous light with a reduction in light, followed by addition of light.

36. The method of claim 35, wherein the reduction in light occurs for a period selected from the group consisting of at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or more than 12 hours.

37. The method of claim 31, wherein the microbes are of the Scendesmus genus and the stimulus is the addition of chromium to the culture media.

38. The method of claim 31, wherein the desired phenotype is hydrogen production.

39. The method of claim 31, wherein nucleic acid exchange occurs between only two parental cells at a time during the mating process.

Figure 1

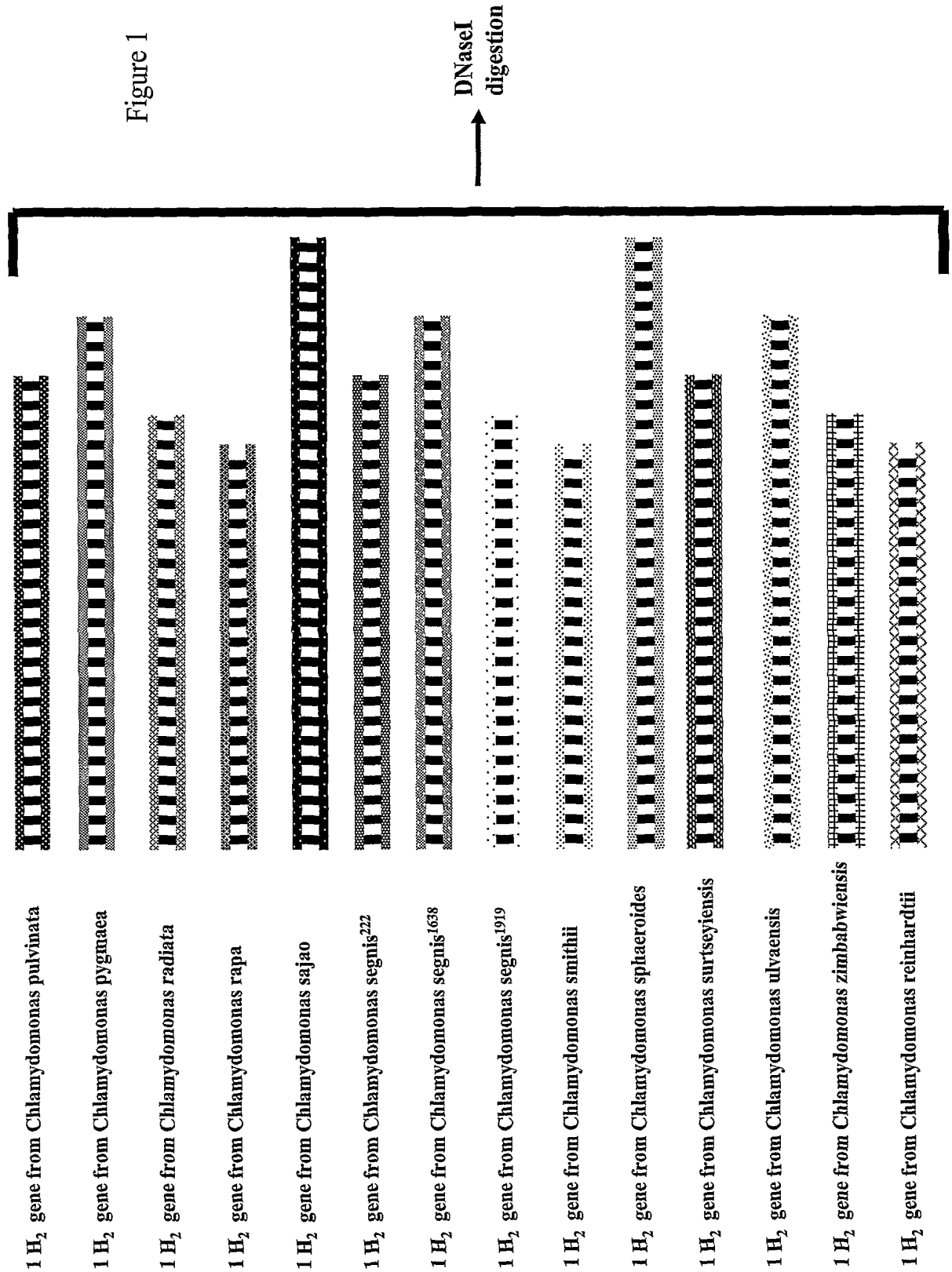
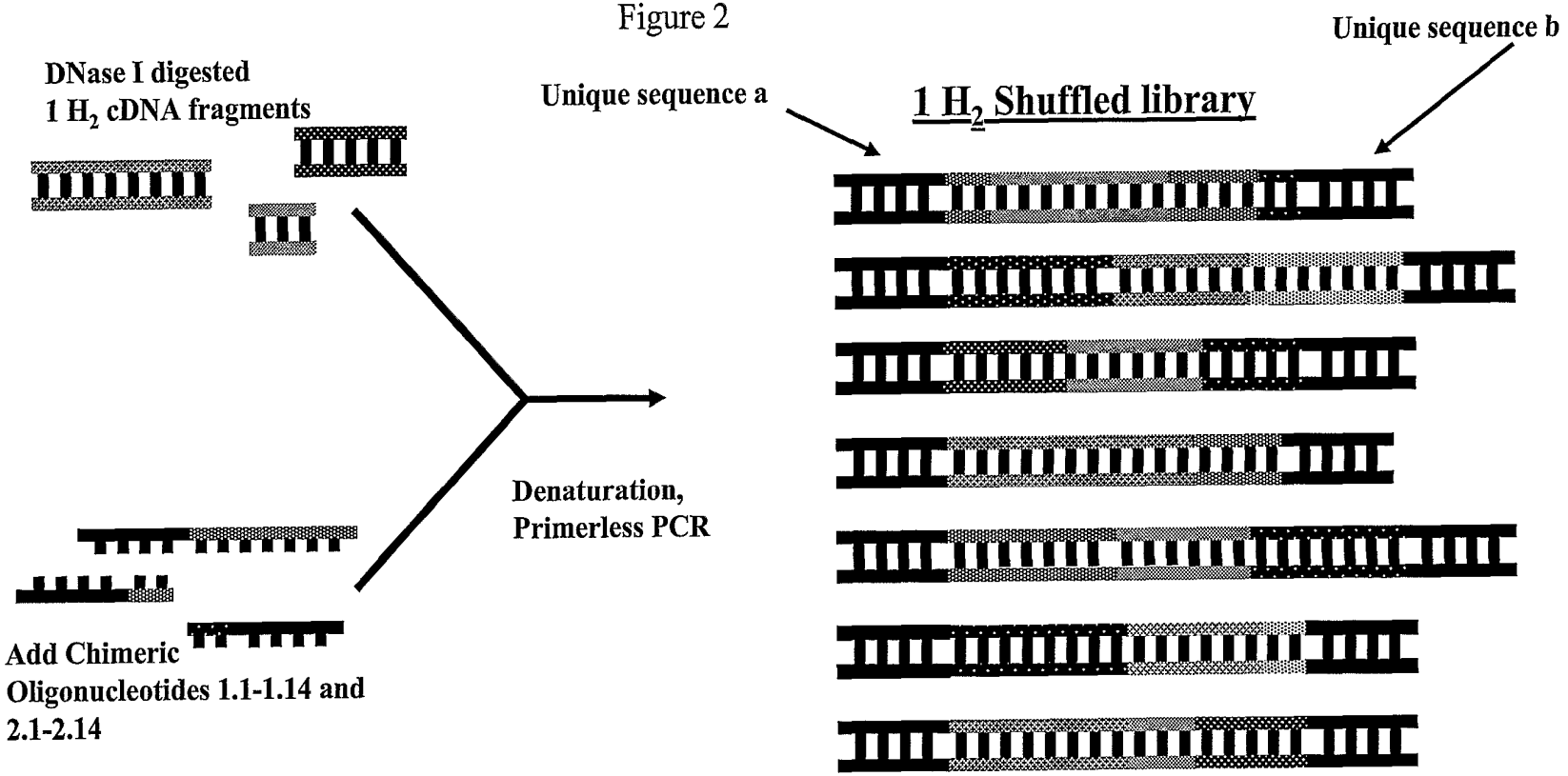

DNaseI digestion
1 H_2 gene from Chlamydomonas pulvinata
1 H_2 gene from Chlamydomonas pygmaea
1 H_2 gene from Chlamydomonas radiata
1 H_2 gene from Chlamydomonas rapa
1 H_2 gene from Chlamydomonas sajao
1 H_2 gene from Chlamydomonas segnis^{222}
1 H_2 gene from Chlamydomonas segnis^{1638}
1 H_2 gene from Chlamydomonas segnis^{1919}
1 H_2 gene from Chlamydomonas smithii
1 H_2 gene from Chlamydomonas sphaeroides
1 H_2 gene from Chlamydomonas surtseyiensis
1 H_2 gene from Chlamydomonas ulvaensis
1 H_2 gene from Chlamydomonas zimbabwiensis
1 H_2 gene from Chlamydomonas reinhardtii

Figure 2



Shuffled $2H_2$, $3H_2$, $4H_2$, and $5H_2$ libraries are created through the same method using chimeric oligonucleotides 3.1-3.14/4.1-4.14, 5.1-5.14/6.1-6.14, 7.1-7.14/8.1-8.14 and 9.1-9.14/10.1-10.14, respectively. Sequences 2 H_2, 3 H_2, 4 H_2, and 5 H_2 are flanked by unique sequences c and d, e and f, g and h, and i and j, respectively.

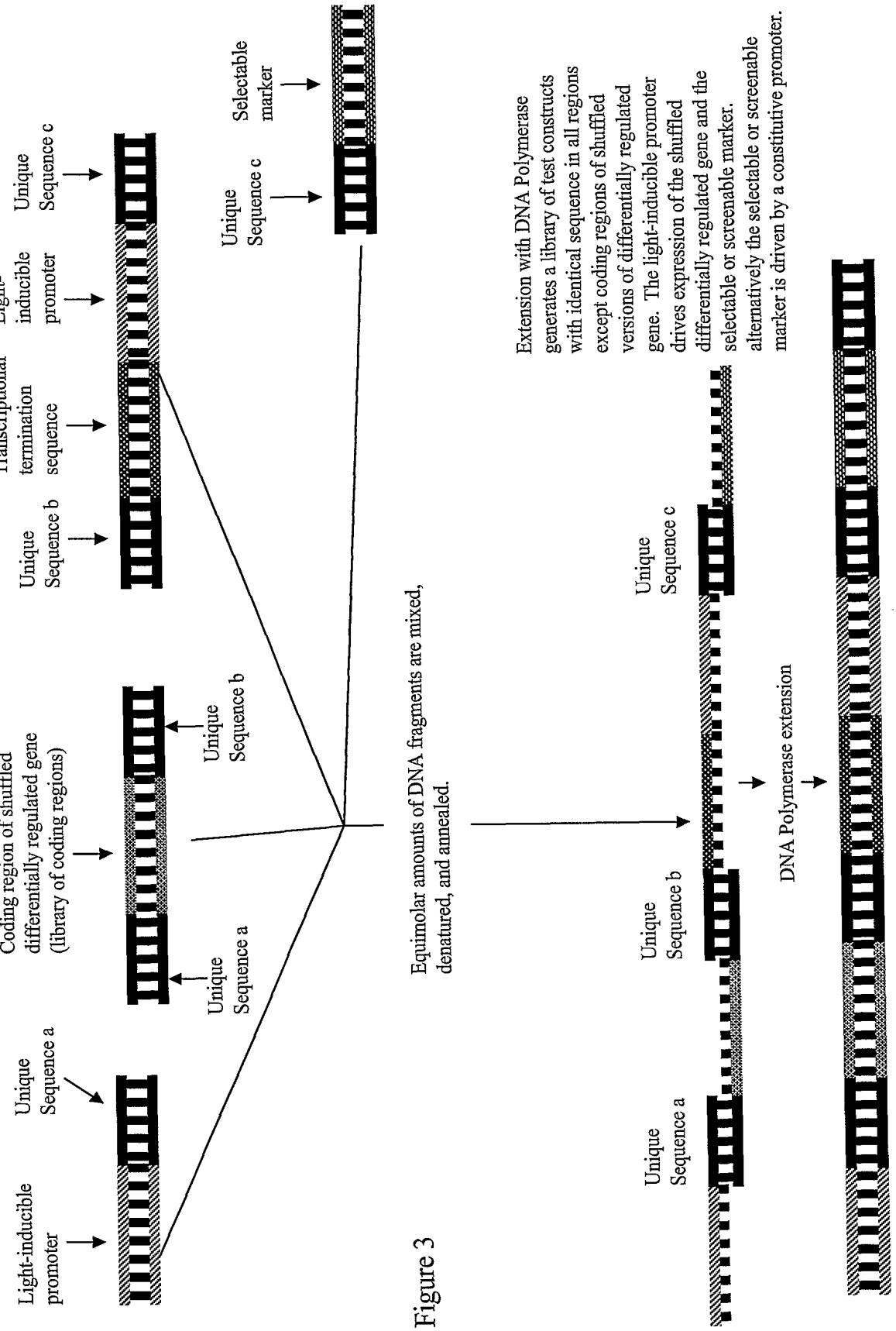

Light-inducible promoter
Unique Sequence a
Coding region of shuffled differentially regulated gene (library of coding regions)
Unique Sequence a
Unique Sequence b
Unique Sequence b
Transcriptional termination sequence
Light-inducible promoter
Unique Sequence c
Unique Sequence c
Selectable marker
Equimolar amounts of DNA fragments are mixed, denatured, and annealed.
Extension with DNA Polymerase generates a library of test constructs with identical sequence in all regions except coding regions of shuffled versions of differentially regulated gene. The light-inducible promoter drives expression of the shuffled differentially regulated gene and the selectable or screenable marker. alternatively the selectable or screenable marker is driven by a constitutive promoter.
Unique Sequence a
Unique Sequence b
Unique Sequence c
DNA Polymerase extension

Figure 3

Figure 4

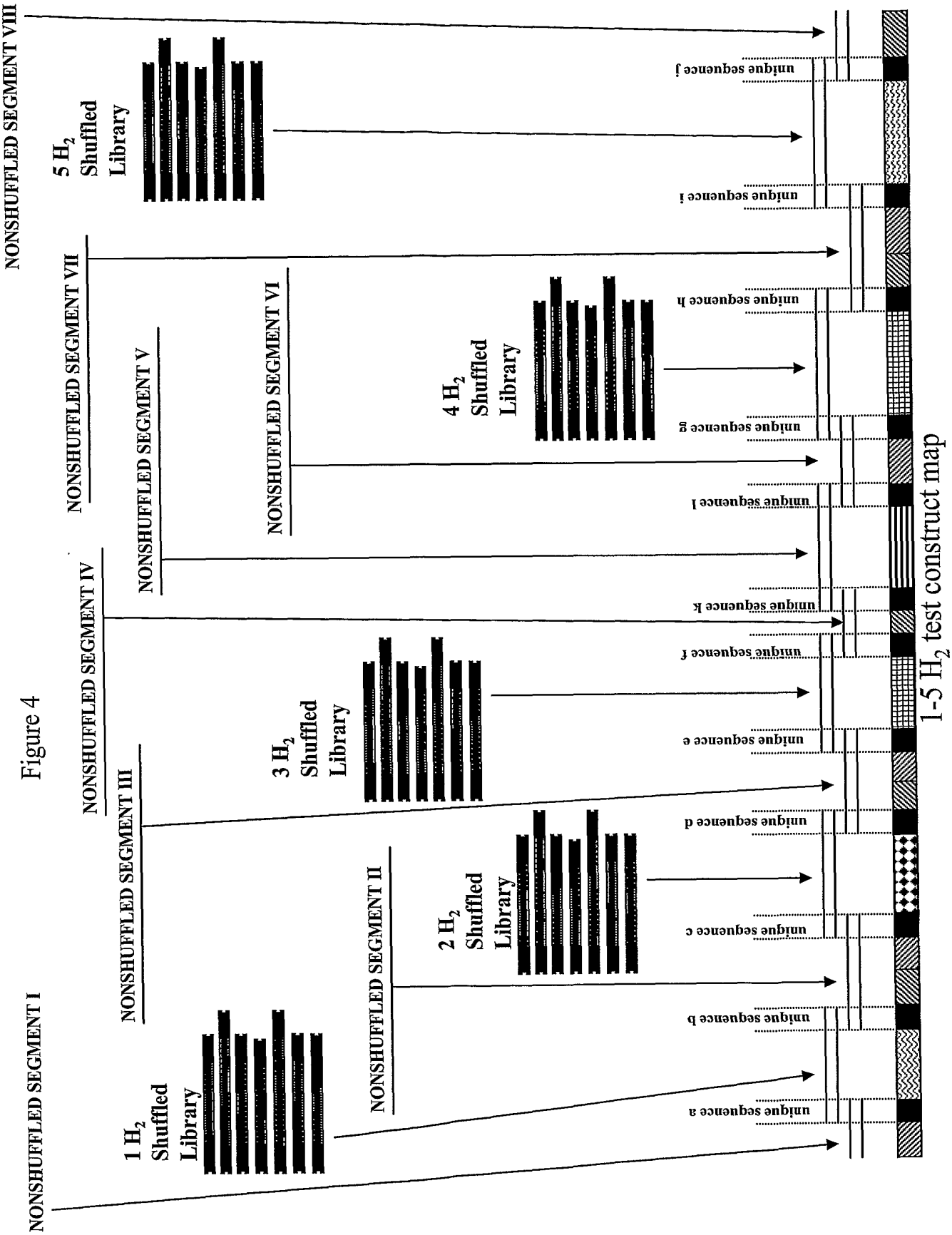
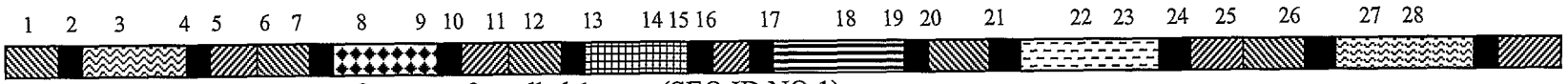
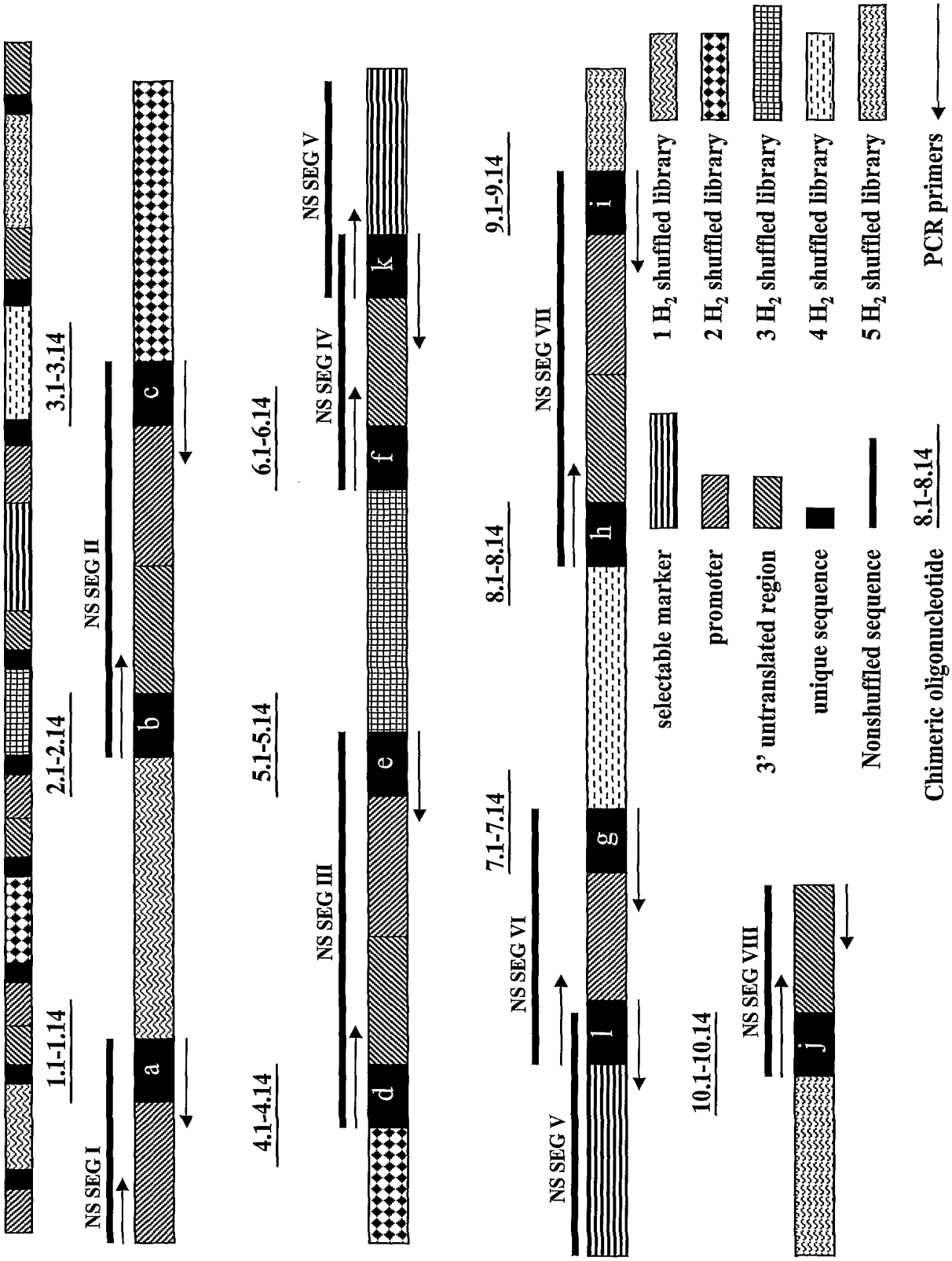

NONSHUFFLED SEGMENT I
NONSHUFFLED SEGMENT II
NONSHUFFLED SEGMENT III
NONSHUFFLED SEGMENT IV
NONSHUFFLED SEGMENT V
NONSHUFFLED SEGMENT VI
NONSHUFFLED SEGMENT VII
NONSHUFFLED SEGMENT VIII
1 H2 Shuffled Library
2 H2 Shuffled Library
3 H2 Shuffled Library
4 H2 Shuffled Library
5 H2 Shuffled Library
unique sequence a
unique sequence b
unique sequence c
unique sequence d
unique sequence e
unique sequence f
unique sequence k
unique sequence l
unique sequence g
unique sequence h
unique sequence i
unique sequence j
1-5 H2 test construct map

Figure 5 1-5 H_2 test constructs



1: light-sensitive promoter fragment from lhcb1 gene (SEQ ID NO 1)
2: unique sequence a (SEQ ID NO 5)
3: shuffled 1 H_2 gene library
4: unique sequence b (SEQ ID NO 6)
5: 3' untranslated region from RBCS2 gene (SEQ ID NO 3)
6: light-sensitive promoter fragment from lhcb1 gene (SEQ ID NO 1)
7: unique sequence c (SEQ ID NO 7)
8: shuffled 2 H_2 gene library
9: unique sequence d (SEQ ID NO 8)
10: 3' untranslated region from RBCS2 gene (SEQ ID NO 3)
11: light-sensitive promoter fragment from lhcb1 gene (SEQ ID NO 1)
12: unique sequence e (SEQ ID NO 9)
13: shuffled 3 H_2 gene library
14: unique sequence f (SEQ ID NO 10)
15: 3' untranslated region from RBCS2 gene (SEQ ID NO 3)
16: unique sequence k (SEQ ID NO 15)
17: ble selectable marker construct (SEQ ID NO 2)
18: unique sequence l (SEQ ID NO 16)
19: light-sensitive promoter fragment from lhcb1 gene (SEQ ID NO 1)
20: unique sequence g (SEQ ID NO 11)
21: shuffled 4 H_2 gene library
22: unique sequence h (SEQ ID NO 12)
23: 3' untranslated region from RBCS2 gene (SEQ ID NO 3)
24: light-sensitive promoter fragment from lhcb1 gene (SEQ ID NO 1)
25: unique sequence i (SEQ ID NO 13)
26: shuffled 5 H_2 gene library
27: unique sequence j (SEQ ID NO 14)
28: 3' untranslated region from RBCS2 gene (SEQ ID NO 3)

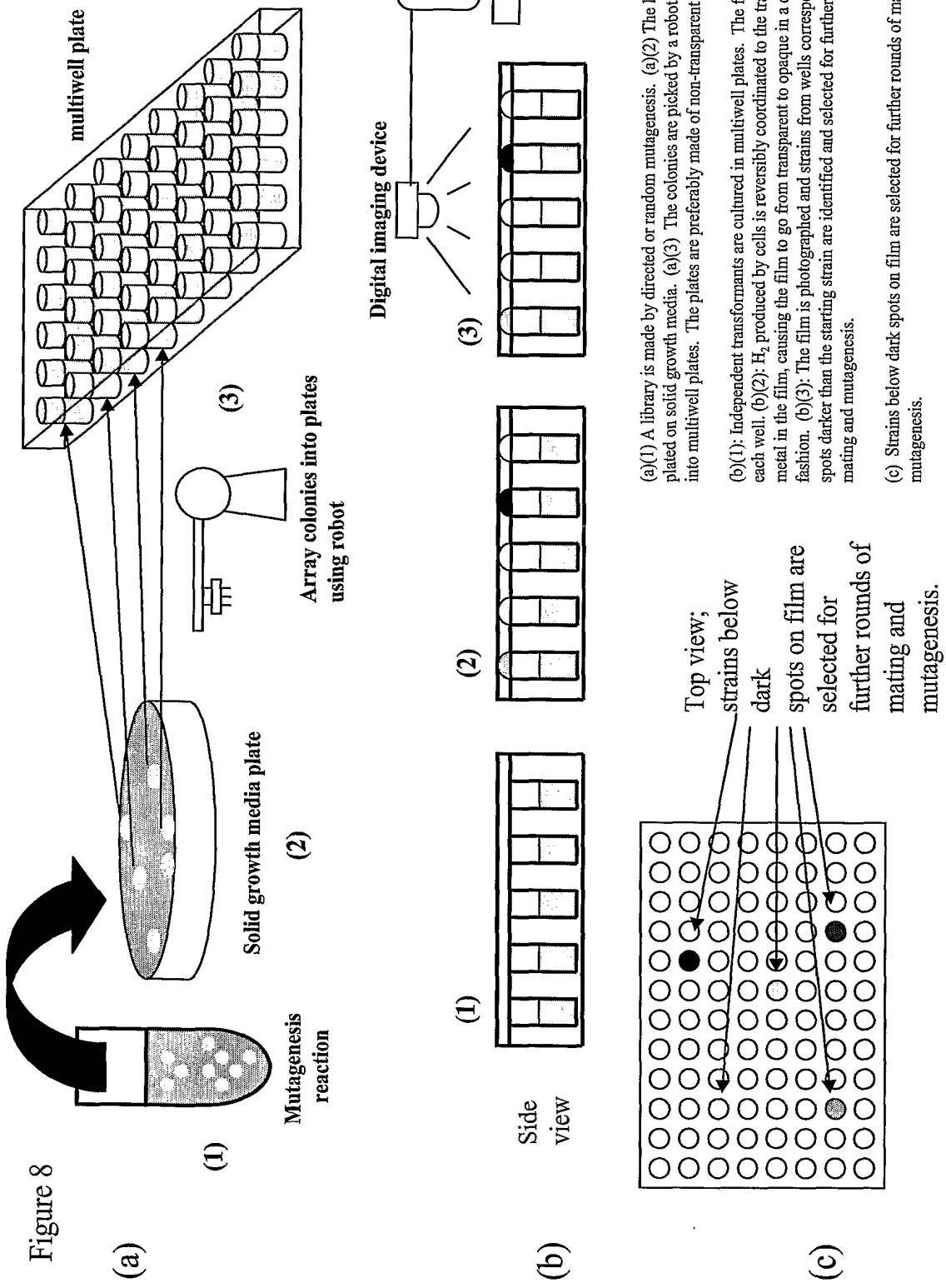
Figure 6


1.1-1.14
2.1-2.14
3.1-3.14
4.1-4.14
5.1-5.14
6.1-6.14
7.1-7.14
8.1-8.14
9.1-9.14
10.1-10.14
NS SEG I
NS SEG II
NS SEG III
NS SEG IV
NS SEG V
NS SEG VI
NS SEG VII
NS SEG VIII
a
b
c
d
e
f
g
h
i
j
k
l
selectable marker
promoter
3' untranslated region
unique sequence
Nonshuffled sequence
Chimeric oligonucleotide
1 H2 shuffled library
2 H2 shuffled library
3 H2 shuffled library
4 H2 shuffled library
5 H2 shuffled library
PCR primers

Hydrogenase-ferredoxin test construct map
11.1-11.6
12.1-12.6
13
NS SEG IX
NS SEG X
NS SEG XI
NS SEG XII
NS SEG XIII
a
b
c
d
e
f
g
h
14
Shuffled hydrogenase library
Ferredoxin library
Green fluorescent protein
13.1-13.6
Chimeric oligonucleotide
Unique sequence
3' untranslated region
promoter
PCR primer

Figure 7

Figure 8



(a)(1) A library is made by directed or random mutagenesis. (a)(2) The library is plated on solid growth media. (a)(3) The colonies are picked by a robot and put into multiwell plates. The plates are preferably made of non-transparent material.

(b)(1): Independent transformants are cultured in multiwell plates. The film seals each well. (b)(2): H_2 produced by cells is reversibly coordinated to the transition metal in the film, causing the film to go from transparent to opaque in a quantitative fashion. (b)(3): The film is photographed and strains from wells corresponding to spots darker than the starting strain are identified and selected for further rounds of mating and mutagenesis.

(c) Strains below dark spots on film are selected for further rounds of mating and mutagenesis.

Figure 9

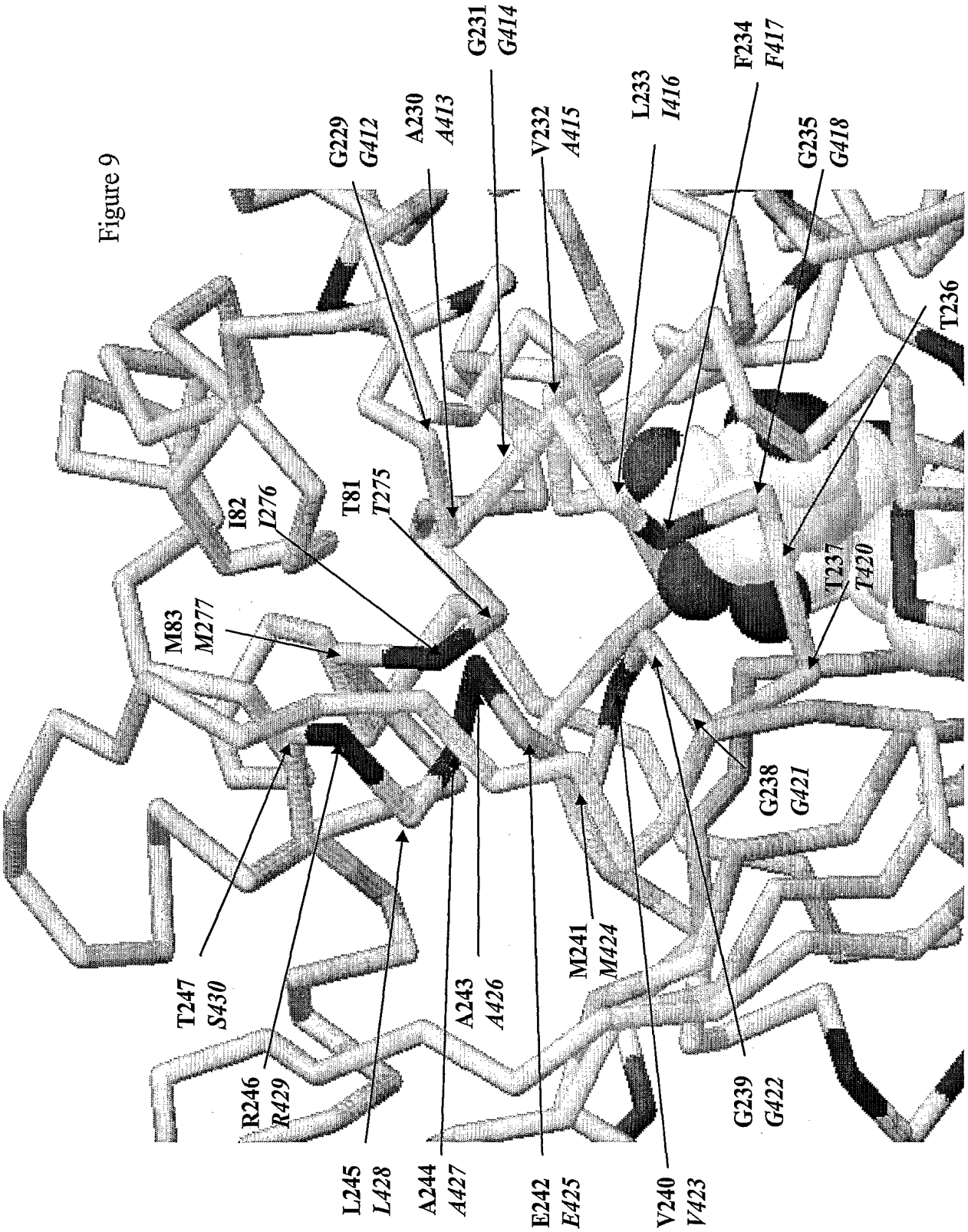

G231
G414
F234
F417
G229
G412
A230
A413
V232
A415
L233
I416
G235
G418
T236
A419
I82
I276
T81
T275
M83
M277
T237
T420
G238
G421
T247
S430
R246
R429
A243
A426
M241
M424
G239
G422
L245
L428
A244
A427
E242
E425
V240
V423

Figure 10

C. Reinhardtii Codon Usage Table: most preferred codons are shown underlined and in bold-face type

TTT 4.8 Phe	TCT 4.7 Ser	TAT 2.6 Tyr	TGT 1.4 Cys
TTC 29.0 Phe	TCC 16.8 Ser	**TAC 23.8 Tyr**	**TGC 12.8 Cys**
TTA 0.7 Leu	TCA 3.0 Ser	**TAA 1.2 STOP**	TGA 0.5 STOP
TTG 3.7 Leu	TCG 16.2 Ser	TAG 0.4 STOP	**TGG 13.5 Trp**
CTT 4.5 Leu	CCT 7.1 Pro	CAT 2.3 His	CGT 5.2 Arg
CTC 12.4 Leu	**CCC 29.9 Pro**	**CAC 17.5 His**	**CGC 35.3 Arg**
CTA 2.4 Leu	CCA 4.4 Pro	CAA 4.2 Gln	CGA 1.8 Arg
CTG 65.4 Leu	CCG 18.6 Pro	**CAG 36.2 Gln**	CGG 9.7 Arg
ATT 9.0 Ile	ACT 5.6 Thr	AAT 2.8 Asn	AGT 2.4 Ser
ATC 28.0 Ile	**ACC 29.9 Thr**	**AAC 29.9 Asn**	**AGC 20.8 Ser**
ATA 0.9 Ile	ACA 3.7 Thr	AAA 2.2 Lys	AGA 0.6 Arg
ATG 26.8 Met	ACG 14.7 Thr	**AAG 46.6 Lys**	AGG 2.5 Arg
GTT 5.3 Val	GCT 17.6 Ala	GAT 7.0 Asp	GGT 10.3 Gly
GTC 16.3 Val	**GCC 55.2 Ala**	**GAC 41.6 Asp**	**GGC 63.2 Gly**
GTA 2.0 Val	GCA 9.6 Ala	GAA 2.5 Glu	GGA 4.9 Gly
GTG 45.6 Val	GCG 38.6 Ala	**GAG 53.4 Glu**	GGG 8.3 Gly

Figure 11

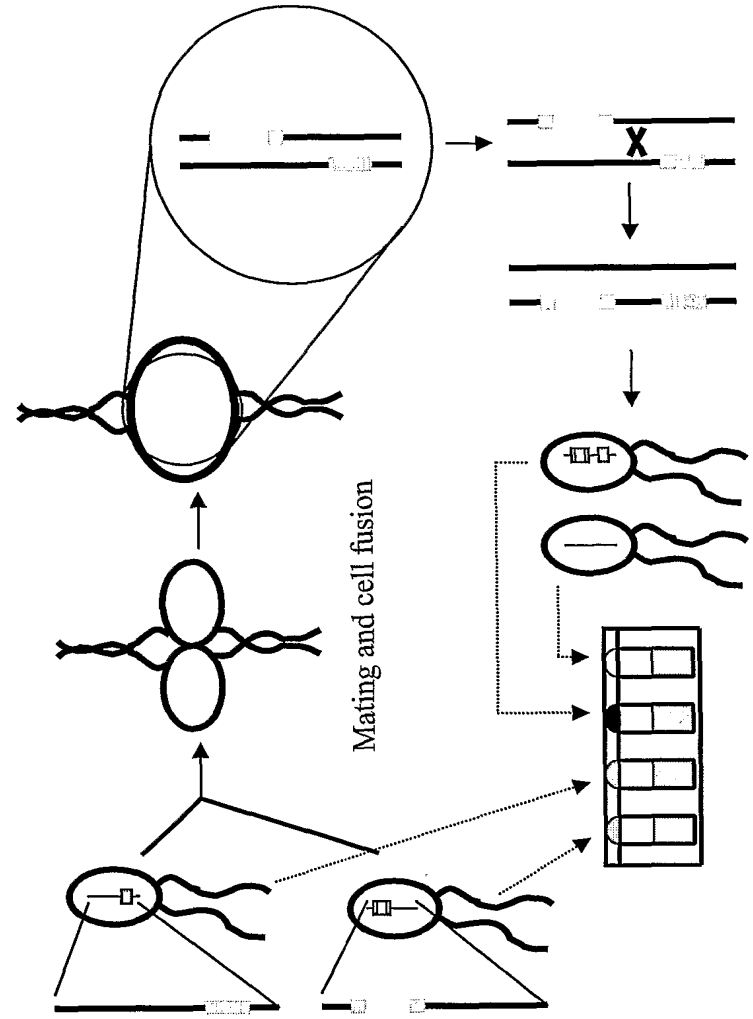

Mating and cell fusion
Parent cell A
Chromosome III: contains
A critical H_2 pathway gene
Parent cell B
Chromosome III: contains
insertion knockout of
deleterious gene

Figure 12

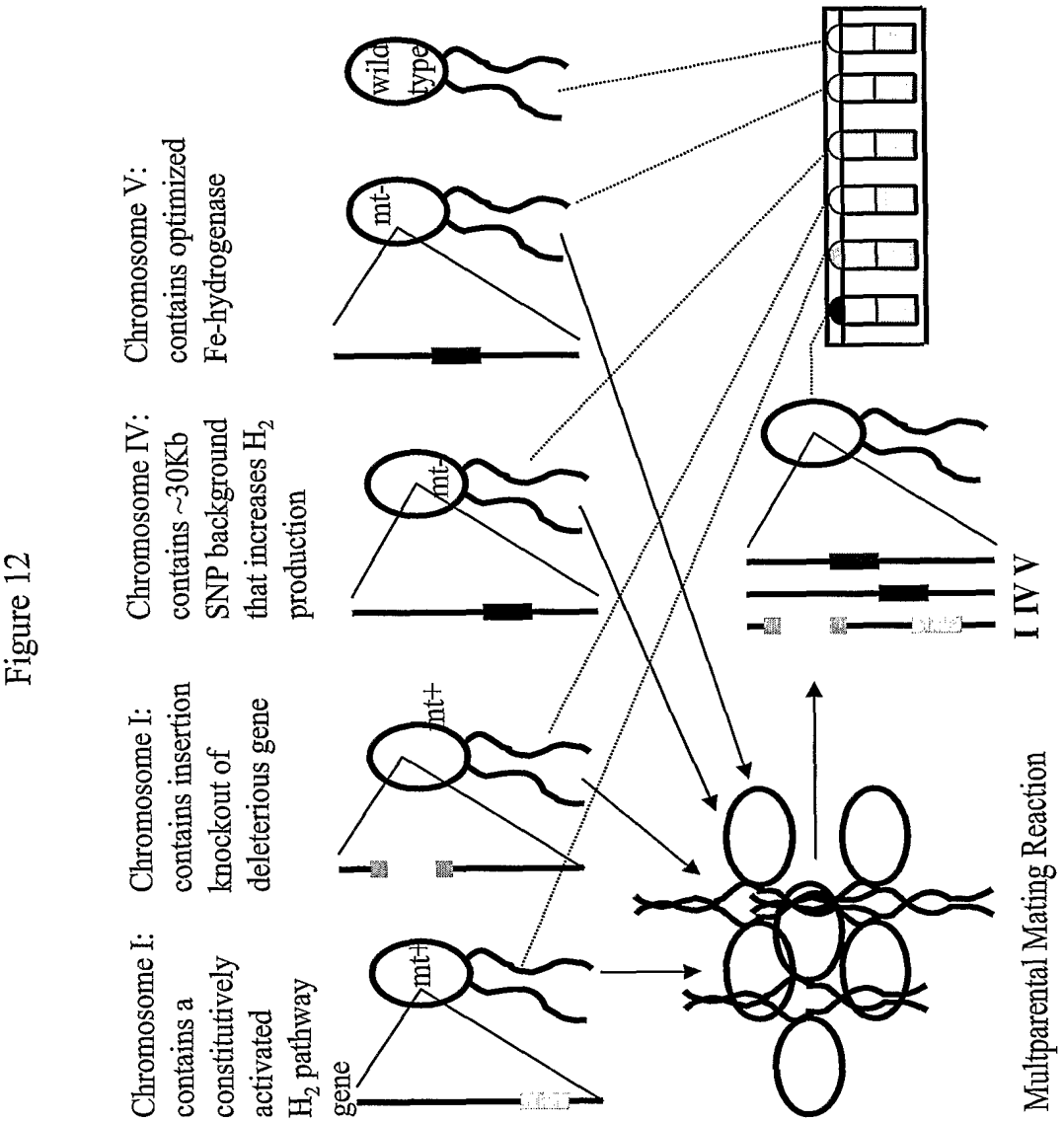

Chromosome I: contains a constitutively activated H_2 pathway gene
Chromosome I: contains insertion knockout of deleterious gene
Chromosome IV: contains ~30Kb SNP background that increases H_2 production
Chromosome V: contains optimized Fe-hydrogenase
mt+
mt+
mt-
wild type
Multparental Mating Reaction
I IV V

Figure 13

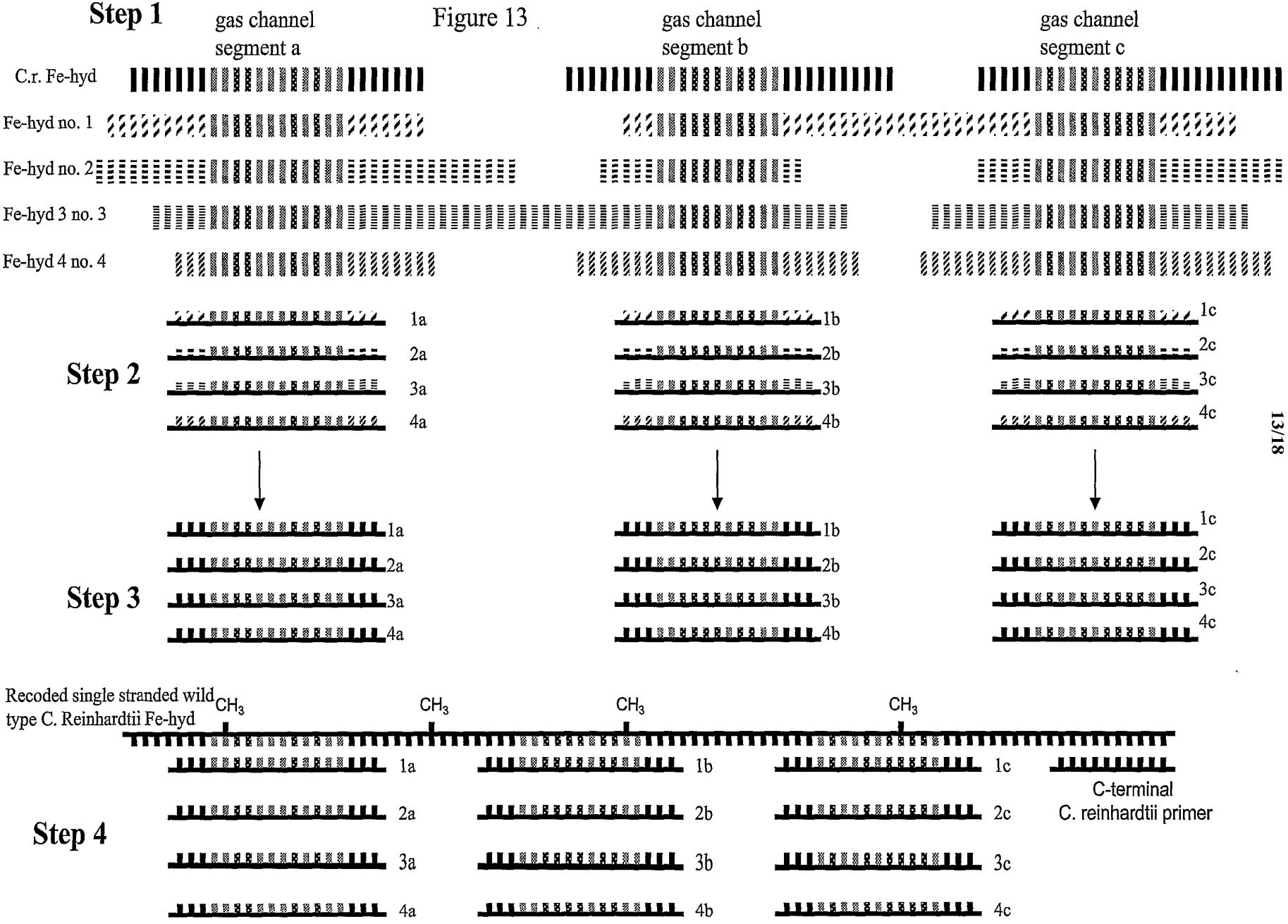

Step 1
gas channel segment a
gas channel segment b
gas channel segment c
C.r. Fe-hyd
Fe-hyd no. 1
Fe-hyd no. 2
Fe-hyd 3 no. 3
Fe-hyd 4 no. 4
Step 2
1a
2a
3a
4a
1b
2b
3b
4b
1c
2c
3c
4c
Step 3
Recoded single stranded wild type C. Reinhardtii Fe-hyd
CH3
Step 4
C-terminal C. reinhardtii primer

Figure 14

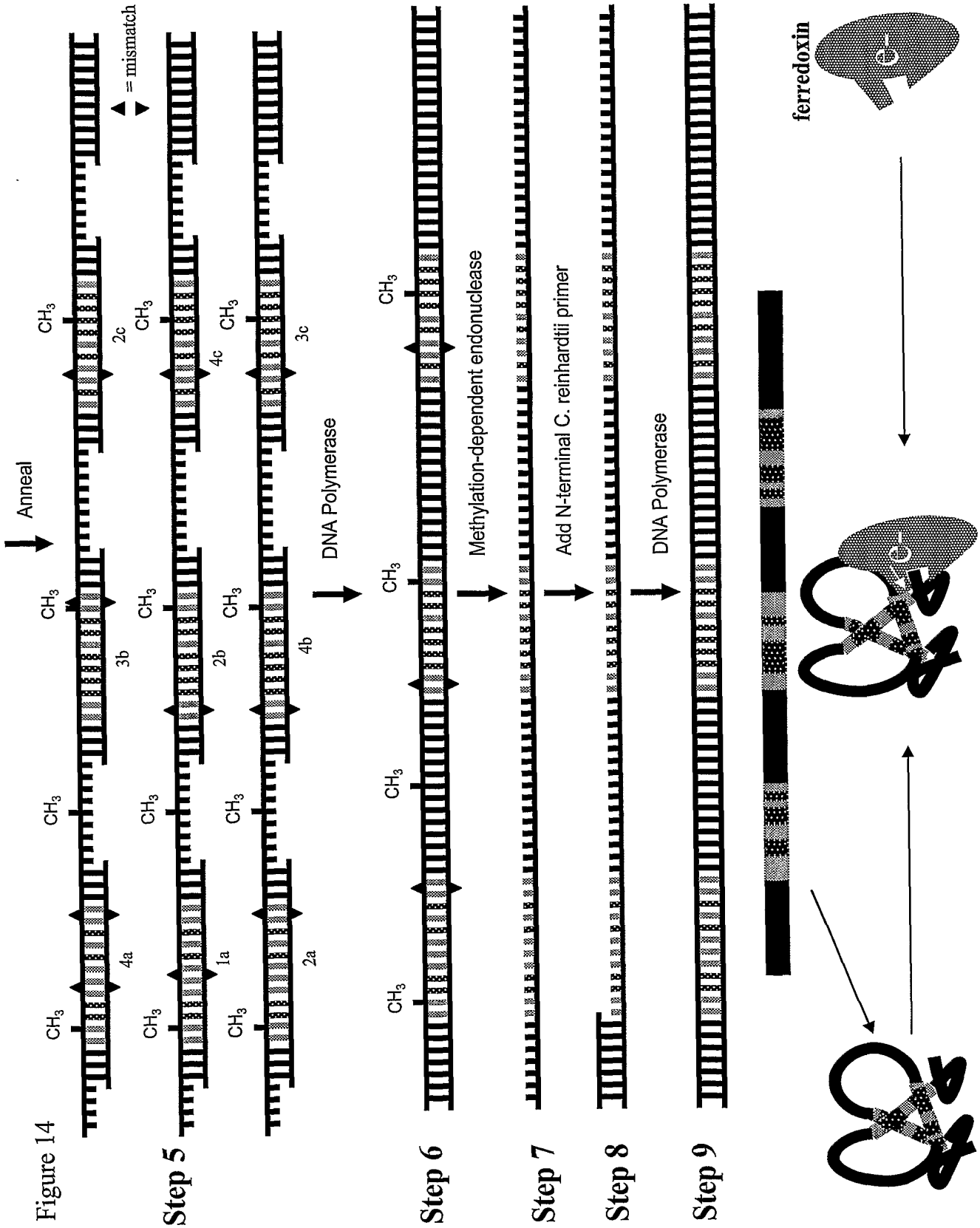

Anneal
= mismatch
CH3
4a
1a
2a
2c
3b
4c
3c
2b
4b
Step 5
DNA Polymerase
Step 6
Methylation-dependent endonuclease
Step 7
Add N-terminal C. reinhardtii primer
Step 8
DNA Polymerase
Step 9
ferredoxin

Figure 15

Residues within 3 amino acids of Fe-hyd most conserved residues (such as $X^1X^2X^3$ of $FX^1X^2X^3GGVMEAA$)
Most highly conserved Fe-hyd residues (such as GGVMEAA)
Framework residues of C. reinhardtii Fe-hydrogenase that are nonconserved in the Fe-hyd gene family
Framework residues of non-C. reinhardtii Fe-hydrogenase no. 1 that are nonconserved in the Fe-hyd gene family
Framework residues of non-C. reinhardtii Fe-hydrogenase no. 2 that are nonconserved in the Fe-hyd gene family
Framework residues of non-C. reinhardtii Fe-hydrogenase no. 3 that are nonconserved in the Fe-hyd gene family
Framework residues of non-C. reinhardtii Fe-hydrogenase no. 4 that are nonconserved in the Fe-hyd gene family

Figure 16

GGVMEAA Aligned Segments from SEQ ID NOs: 1-122: (corresponding to SEQ ID Nos: 124-141)

1. GAGVIFGATGGVMEAALRT
2. GGGAIFCATGGVMEAAVRS
3. GGATIFGVTGGVMEAALRF
4. GAGAIFGATGGVMEAALRS
5. GAGAIFGATGGVMEAAIRS
6. GAAVIFGVTGGVMEAALRT
7. GAGQIFAATGGVMEAASRT
8. GGGVLFGTTGGVMEAALRT
9. GAAVIFGTTGGVMEAALRT
10. GAAPIFGVTGGVIEAALRT
11. GAGVIFGTTGGVMEAALRS
12. GAGVIFGATGGVMEAAIRT
13. SAGNLFGVTGGVMEAAIRT
14. GAGAIFGATGGVMEAALRT
15. GAGVLFGTTGGVMEAALRT
16. GAAALFGVTGGVMEAALRT
17. GAGVLFGTTGGVMEAAVRT
18. GAGTIFGTTGGVMEAALRT

TIMEE Aligned Segments from SEQ ID NOs: 1-122: (corresponding to SEQ ID Nos: 142-147)

1. TIMEE
2. TIVEE
3. TIWEE
4. TICEE
5. VIMEE
6. TARLE

Figure 17

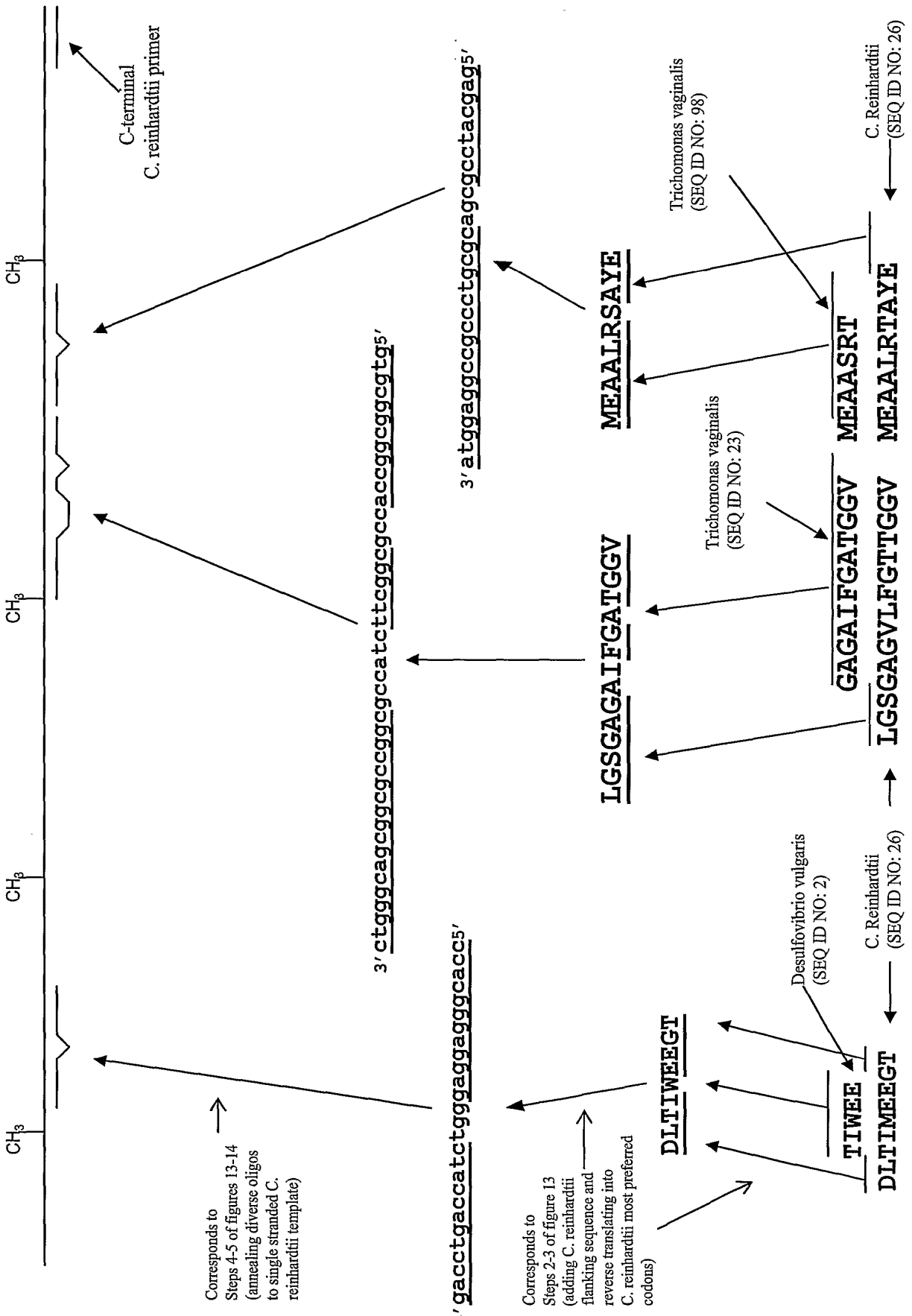
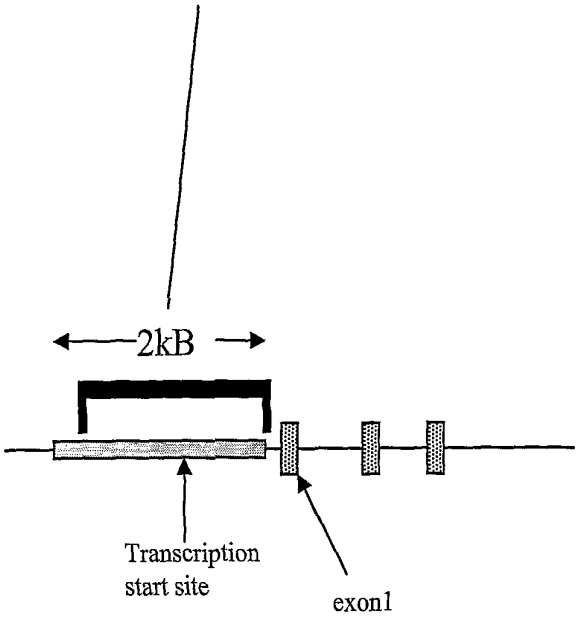
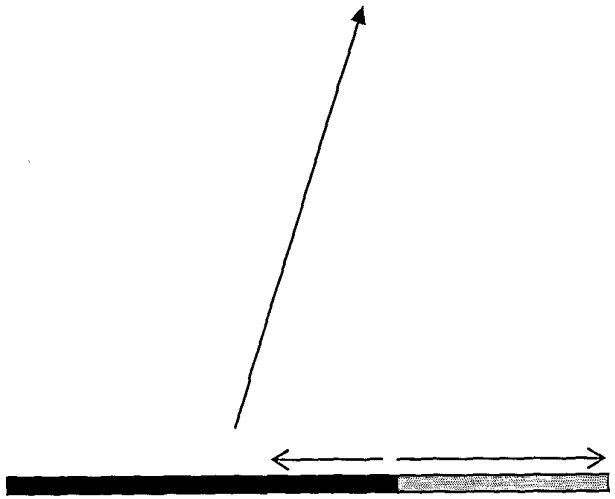

C-terminal
C. reinhardtii primer
CH3
CH3
CH3
3′ctgggcagcgggcgccgggcgccatcttcggcgccaccgggcgcgtg5′
3′atggaggccgccgccctgcgcagcgcctacgag5′
3′gacctgaccatctgggaggagggcaccc5′
LGSGAGAIFGATGGV
MEAALRSAYE
GAGAIFGATGGV
MEAASRT
LGSGAGVLFGTTGGV
MEAALRTAYE
DLTIWEEGT
TIWEE
DLTIMEEGT
Corresponds to Steps 4-5 of figures 13-14 (annealing diverse oligos to single stranded C. reinhardtii template)
Corresponds to Steps 2-3 of figure 13 (adding C. reinhardtii flanking sequence and reverse translating into C. reinhardtii most preferred codons)
Desulfovibrio vulgaris (SEQ ID NO: 2)
C. Reinhardtii (SEQ ID NO: 26)
Trichomonas vaginalis (SEQ ID NO: 23)
Trichomonas vaginalis (SEQ ID NO: 98)
C. Reinhardtii (SEQ ID NO: 26)

Figure 18

Amplify promoters from genes in nuclear, chloroplast, and mitochondrial genomes



Make insertion library with diverse promoters and screen or select for desired phenotype that arises due to altered transcriptional regulation of metabolic pathway.



Selectable marker

Connect to selectable marker through chimeric oligonucleotides and DNA Polymerase extension (marker Gene is transcribed in opposite direction)

050118 CIP Sequence Listing
SEQUENCE LISTING

<110> Solazyme, Inc.
Dillon, Harrison F.

<120> Methods and Compositions for Evolving Microbial Hydrogen Production

<130> H2042101-CIP

<140> US 10/763,712
<141> 2004-01-21

<150> US 10/287,750
<151> 2002-11-04

<150> US 10/411,910
<151> 2003-04-12

<150> US 60/500,032
<151> 2003-09-03

<160> 184

<170> PatentIn version 3.2

<210> 1
<211> 574
<212> PRT
<213> Clostriduim pasteuranum

<400> 1

Met Lys Thr Ile Ile Ile Asn Gly Val Gln Phe Asn Thr Asp Glu Asp
1 5 10 15

Thr Thr Ile Leu Lys Phe Ala Arg Asp Asn Asn Ile Asp Ile Ser Ala
20 25 30

Leu Cys Phe Leu Asn Asn Cys Asn Asn Asp Ile Asn Lys Cys Glu Ile
35 40 45

Cys Thr Val Glu Val Glu Gly Thr Gly Leu Val Thr Ala Cys Asp Thr
50 55 60

Leu Ile Glu Asp Gly Met Ile Ile Asn Thr Asn Ser Asp Ala Val Asn
65 70 75 80

Glu Lys Ile Lys Ser Arg Ile Ser Gln Leu Leu Asp Ile His Glu Phe
85 90 95

Lys Cys Gly Pro Cys Asn Arg Arg Glu Asn Cys Glu Phe Leu Lys Leu
100 105 110

Val Ile Lys Tyr Lys Ala Arg Ala Ser Lys Pro Phe Leu Pro Lys Asp
115 120 125

Lys Thr Glu Tyr Val Asp Glu Arg Ser Lys Ser Leu Thr Val Asp Arg
130 135 140

Thr Lys Cys Leu Leu Cys Gly Arg Cys Val Asn Ala Cys Gly Lys Asn
145 150 155 160

Thr Glu Thr Tyr Ala Met Lys Phe Leu Asn Lys Asn Gly Lys Thr Ile
165 170 175

Ile Gly Ala Glu Asp Glu Lys Cys Phe Asp Asp Thr Asn Cys Leu Leu
180 185 190

Cys Gly Gln Cys Ile Ile Ala Cys Pro Val Ala Ala Leu Ser Glu Lys
195 200 205

Ser His Met Asp Arg Val Lys Asn Ala Leu Asn Ala Pro Glu Lys His
210 215 220

Val Ile Val Ala Met Ala Pro Ser Val Arg Ala Ser Ile Gly Glu Leu
225 230 235 240

Phe Asn Met Gly Phe Gly Val Asp Val Thr Gly Lys Ile Tyr Thr Ala
245 250 255

Leu Arg Gln Leu Gly Phe Asp Lys Ile Phe Asp Ile Asn Phe Gly Ala
260 265 270

Asp Met Thr Ile Met Glu Glu Ala Thr Glu Leu Val Gln Arg Ile Glu
275 280 285

Asn Asn Gly Pro Phe Pro Met Phe Thr Ser Cys Cys Pro Gly Trp Val
290 295 300

Arg Gln Ala Glu Asn Tyr Tyr Pro Glu Leu Leu Asn Asn Leu Ser Ser
305 310 315 320

Ala Lys Ser Pro Gln Gln Ile Phe Gly Thr Ala Ser Lys Thr Tyr Tyr
325 330 335

Pro Ser Ile Ser Gly Leu Asp Pro Lys Asn Val Phe Thr Val Thr Val
340 345 350

Met Pro Cys Thr Ser Lys Lys Phe Glu Ala Asp Arg Pro Gln Met Glu
355 360 365

Lys Asp Gly Leu Arg Asp Ile Asp Ala Val Ile Thr Thr Arg Glu Leu
370 375 380

Ala Lys Met Ile Lys Asp Ala Lys Ile Pro Phe Ala Lys Leu Glu Asp
385 390 395 400

Ser Glu Ala Asp Pro Ala Met Gly Glu Tyr Ser Gly Ala Gly Ala Ile
405 410 415

Phe Gly Ala Thr Gly Gly Val Met Glu Ala Ala Leu Arg Ser Ala Lys
420 425 430

Asp Phe Ala Glu Asn Ala Glu Leu Glu Asp Ile Glu Tyr Lys Gln Val
435 440 445

Arg Gly Leu Asn Gly Ile Lys Glu Ala Glu Val Glu Ile Asn Asn Asn
450 455 460

Lys Tyr Asn Val Ala Val Ile Asn Gly Ala Ser Asn Leu Phe Lys Phe

475 480

Met Lys Ser Gly Met Ile Asn Glu Lys Gln Tyr His Phe Ile Glu Val
485 490 495

Met Ala Cys His Gly Gly Cys Val Asn Gly Gly Gly Gln Pro His Val
500 505 510

Asn Pro Lys Asp Leu Glu Lys Val Asp Ile Lys Lys Val Arg Ala Ser
515 520 525

Val Leu Tyr Asn Gln Asp Glu His Leu Ser Lys Arg Lys Ser His Glu
530 535 540

Asn Thr Ala Leu Val Lys Met Tyr Gln Asn Tyr Phe Gly Lys Pro Gly
545 550 555 560

Glu Gly Arg Ala His Glu Ile Leu His Phe Lys Tyr Lys Lys
565 570

<210> 2
<211> 421
<212> PRT
<213> Desulfovibrio vulgaris

<400> 2

Met Ser Arg Thr Val Met Glu Arg Ile Glu Tyr Glu Met His Thr Pro
1 5 10 15

Asp Pro Lys Ala Asp Pro Asp Lys Leu His Phe Val Gln Ile Asp Glu
20 25 30

Ala Lys Cys Ile Gly Cys Asp Thr Cys Ser Gln Tyr Cys Pro Thr Ala
35 40 45

Ala Ile Phe Gly Glu Met Gly Glu Pro His Ser Ile Pro His Ile Glu
50 55 60

Ala Cys Ile Asn Cys Gly Gln Cys Leu Thr His Cys Pro Glu Asn Ala
65 70 75 80

Ile Tyr Glu Ala Gln Ser Trp Val Pro Glu Val Glu Lys Lys Leu Lys
85 90 95

Asp Gly Lys Val Lys Cys Ile Ala Met Pro Ala Pro Ala Val Arg Tyr
100 105 110

Ala Leu Gly Asp Ala Phe Gly Met Pro Val Gly Ser Val Thr Thr Gly
115 120 125

Lys Met Leu Ala Ala Leu Gln Lys Leu Gly Phe Ala His Cys Trp Asp
130 135 140

Thr Glu Phe Thr Ala Asp Val Thr Ile Trp Glu Glu Gly Ser Glu Phe
145 150 155 160

Val Glu Arg Leu Thr Lys Lys Ser Asp Met Pro Leu Pro Gln Phe Thr

170 175

Ser Cys Cys Pro Gly Trp Gln Lys Tyr Ala Glu Thr Tyr Tyr Pro Glu
180 185 190

Leu Leu Pro His Phe Ser Thr Cys Lys Ser Pro Ile Gly Met Asn Gly
195 200 205

Ala Leu Ala Lys Thr Tyr Gly Ala Glu Arg Met Lys Tyr Asp Pro Lys
210 215 220

Gln Val Tyr Thr Val Ser Ile Met Pro Cys Ile Ala Lys Lys Tyr Glu
225 230 235 240

Gly Leu Arg Pro Glu Leu Lys Ser Ser Gly Met Arg Asp Ile Asp Ala
245 250 255

Thr Leu Thr Thr Arg Glu Leu Ala Tyr Met Ile Lys Lys Ala Gly Ile
260 265 270

Asp Phe Ala Lys Leu Pro Asp Gly Lys Arg Asp Ser Leu Met Gly Glu
275 280 285

Ser Thr Gly Gly Ala Thr Ile Phe Gly Val Thr Gly Gly Val Met Glu
290 295 300

Ala Ala Leu Arg Phe Ala Tyr Glu Ala Val Thr Gly Lys Lys Pro Asp
305 310 315 320

Ser Trp Asp Phe Lys Ala Val Arg Gly Leu Asp Gly Ile Lys Glu Ala
325 330 335

Thr Val Asn Val Gly Gly Thr Asp Val Lys Val Ala Val Val His Gly
340 345 350

Ala Lys Arg Phe Lys Gln Val Cys Asp Asp Val Lys Ala Gly Lys Ser
355 360 365

Pro Tyr His Phe Ile Glu Tyr Met Ala Cys Pro Gly Gly Cys Val Cys
370 375 380

Gly Gly Gly Gln Pro Val Met Pro Gly Val Leu Glu Ala Met Asp Arg
385 390 395 400

Thr Thr Thr Arg Leu Tyr Ala Gly Leu Lys Lys Arg Leu Ala Met Ala
405 410 415

Ser Ala Asn Lys Ala
420

<210> 3
<211> 468
<212> PRT
<213> Entamoeba histolytica

<400> 3

Met Pro Pro Lys Pro Ser His Thr Leu Thr Gly His Asp His Asn His

```
                                                    10                  15

Ser Ile Gln Phe Asp Trp Ser Lys Cys Met Gly Cys Gly Met Cys Ala
            20                  25                  30

Thr Lys Cys Thr Phe Gly Val Leu Val Lys Gln Pro Pro Lys Ile Pro
        35                  40                  45

Pro Phe Val Gln Pro Asn Arg Glu Lys Leu Ser Gln Glu Asn Thr Asp
    50                  55                  60

Lys Thr Arg Val Leu Ile Asp Glu Ser Glu Cys Thr Gly Cys Gly Gln
65                  70                  75                  80

Cys Ser Leu Val Cys Asn Phe Gly Ser Ile Thr Pro Ile Asp His Leu
                85                  90                  95

Val Asp Thr Phe Lys Ala Lys Glu Ala Gly Lys Lys Leu Val Ala Met
            100                 105                 110

Ile Ala Pro Ser Thr Arg Leu Gly Val Ala Glu Ala Met Gly Met Pro
        115                 120                 125

Ile Gly Ser Thr Ala Met Ala Gln Leu Val His Cys Leu Arg Leu Ile
    130                 135                 140

Gly Phe Asp Tyr Val Phe Asp Val Asp Ala Gly Ala Asp Lys Thr Thr
145                 150                 155                 160

Met Asp Asp Tyr Ala Glu Val Ile Glu Met Lys Lys Glu Gly Lys Gly
                165                 170                 175

Pro Ala Ile Thr Ser Cys Cys Pro Ala Trp Ile Glu Leu Val Glu Lys
            180                 185                 190

Glu Tyr Pro Asp Leu Ile Pro Asn Val Ser Thr Ala Arg Ser Pro Ile
        195                 200                 205

Gly Cys Leu Ala Gly Cys Ile Lys Arg Gly Trp Ala Lys Asp Val Gly
    210                 215                 220

Ile Ala Val Glu Asp Leu Tyr Thr Val Gly Ile Met Pro Cys Ile Ala
225                 230                 235                 240

Lys Lys Thr Glu Ser Gln Arg Gln Gln Ile His Gln Asp Tyr Asp Ala
                245                 250                 255

Ser Cys Thr Ser Asn Glu Ile Ala Ala Tyr Phe Lys Lys His Leu Pro
            260                 265                 270

Pro Glu Glu Cys Lys Phe Thr Gln Glu Arg Glu Glu Ala Leu Ala Lys
        275                 280                 285

Thr Glu Asp Gly Gln Cys Asp Leu Pro Phe Arg Arg Ile Ser Gly Gly
    290                 295                 300
```

Ser Asn Ile Phe Gly Lys Thr Gly Gly Val Cys Glu Thr Val Leu Arg
305 310 315 320

Val Ile Ala Arg Asn Ala Gly Val Asp Trp Asn Ser Cys Thr Val Asn
325 330 335

Lys Glu Glu Thr Phe Lys His Ala Ala Ser Gly Ser Thr Met Thr Asn
340 345 350

Leu Ser Val Asp Ile Gly Gly Thr Ile Ile Thr Gly Ala Val Cys His
355 360 365

Gly Gly Tyr Ala Ile Arg His Ala Cys Glu Leu Ile Arg Lys Gly Glu
370 375 380

Leu Lys Val Asp Val Val Glu Met Met Ala Cys Val Gly Gly Cys Leu
385 390 395 400

Gly Gly Ala Gly Gln Pro Lys Ile Pro Pro Ala Lys Lys Leu Glu Met
405 410 415

Asp Lys Arg Arg Val Met Leu Asp Ile Leu Asp Gln Gln Thr Asp Ile
420 425 430

Arg Ala Ala Asn Glu Asn Thr Asp Val Leu Gly Trp Ile Asp Lys His
435 440 445

Phe Asp His Gln Gly Ala His Gln His Leu His Thr Tyr Phe Thr Pro
450 455 460

Arg Tyr Gln Asn
465

<210> 4
<211> 491
<212> PRT
<213> Saccharomyces cerevisiae

<400> 4

Met Ser Ala Leu Leu Ser Glu Ser Asp Leu Asn Asp Phe Ile Ser Pro
1 5 10 15

Ala Leu Ala Cys Val Lys Pro Thr Gln Val Ser Gly Gly Lys Lys Asp
20 25 30

Asn Val Asn Met Asn Gly Glu Tyr Glu Val Ser Thr Glu Pro Asp Gln
35 40 45

Leu Glu Lys Val Ser Ile Thr Leu Ser Asp Cys Leu Ala Cys Ser Gly
50 55 60

Cys Ile Thr Ser Ser Glu Glu Ile Leu Leu Ser Ser Gln Ser His Ser
65 70 75 80

Val Phe Leu Lys Asn Trp Gly Lys Leu Ser Gln Gln Gln Asp Lys Phe
85 90 95

Leu Val Val Ser Val Ser Pro Gln Cys Arg Leu Ser Leu Ala Gln Tyr
100 105 110

Tyr Gly Leu Thr Leu Glu Ala Ala Asp Leu Cys Leu Met Asn Phe Phe
115 120 125

Gln Lys His Phe Gln Cys Lys Tyr Met Val Gly Thr Glu Met Gly Arg
130 135 140

Ile Ile Ser Ile Ser Lys Thr Val Glu Lys Ile Ile Ala His Lys Lys
145 150 155 160

Gln Lys Glu Asn Thr Gly Ala Asp Arg Lys Pro Leu Leu Ser Ala Val
165 170 175

Cys Pro Gly Phe Leu Ile Tyr Thr Glu Lys Thr Lys Pro Gln Leu Val
180 185 190

Pro Met Leu Leu Asn Val Lys Ser Pro Gln Gln Ile Thr Gly Ser Leu
195 200 205

Ile Arg Ala Thr Phe Glu Ser Leu Ala Ile Ala Arg Glu Ser Phe Tyr
210 215 220

His Leu Ser Leu Met Pro Cys Phe Asp Lys Lys Leu Glu Ala Ser Arg
225 230 235 240

Pro Glu Ser Leu Asp Asp Gly Ile Asp Cys Val Ile Thr Pro Arg Glu
245 250 255

Ile Val Thr Met Leu Gln Glu Leu Asn Leu Asp Phe Lys Ser Phe Leu
260 265 270

Thr Glu Asp Thr Ser Leu Tyr Gly Arg Leu Ser Pro Pro Gly Trp Asp
275 280 285

Pro Arg Val His Trp Ala Ser Asn Leu Gly Gly Thr Cys Gly Gly Tyr
290 295 300

Ala Tyr Gln Tyr Val Thr Ala Val Gln Arg Leu His Pro Gly Ser Gln
305 310 315 320

Met Ile Val Leu Glu Gly Arg Asn Ser Asp Ile Val Glu Tyr Arg Leu
325 330 335

Leu His Asp Asp Arg Ile Ile Ala Ala Ala Ser Glu Leu Ser Gly Phe
340 345 350

Arg Asn Ile Gln Asn Leu Val Arg Lys Leu Thr Ser Gly Ser Gly Ser
355 360 365

Glu Arg Lys Arg Asn Ile Thr Ala Leu Arg Lys Arg Arg Thr Gly Pro
370 375 380

Lys Ala Asn Ser Arg Glu Met Ala Ala Ala Thr Ala Ala Thr Ala Asp
385 390 395 400

Pro Tyr His Ser Asp Tyr Ile Glu Val Asn Ala Cys Pro Gly Ala Cys
405 410 415

Met Asn Gly Gly Gly Leu Leu Asn Gly Glu Gln Asn Ser Leu Lys Arg
420 425 430

Lys Gln Leu Val Gln Thr Leu Asn Lys Arg His Gly Glu Glu Leu Ala
435 440 445

Met Val Asp Pro Leu Thr Leu Gly Pro Lys Leu Glu Glu Ala Ala Ala
450 455 460

Arg Pro Leu Ser Leu Glu Tyr Val Phe Ala Pro Val Lys Gln Ala Val
465 470 475 480

Glu Lys Asp Leu Val Ser Val Gly Ser Thr Trp
485 490

<210> 5
<211> 436
<212> PRT
<213> Chlorella fusca

<400> 5

Met Cys Cys Pro Val Val Ala Ser Arg His Ala Gly Arg Ala Arg His
1 5 10 15

Val Ala Val Arg Ala Ala Gly Pro Thr Ser Glu Cys Asp Cys Pro Pro
20 25 30

Thr Pro Gln Ala Lys Leu Pro His Trp Gln Gln Ala Leu Asp Glu Leu
35 40 45

Ala Lys Pro Lys Glu Ser Arg Arg Leu Met Ile Ala Gln Ile Ala Ser
50 55 60

Ala Val Arg Val Ala Ile Ala Glu Thr Ile Gly Leu Ala Pro Gly Asp
65 70 75 80

Val Thr Ile Gly Gln Leu Val Thr Gly Leu Arg Met Leu Gly Phe Asp
85 90 95

Tyr Val Phe Asp Thr Leu Phe Gly Ala Asp Leu Thr Ile Met Glu Glu
100 105 110

Gly Thr Glu Leu Leu His Arg Leu Gln Asp His Leu Glu Gln His Pro
115 120 125

Asn Lys Glu Glu Pro Leu Pro Met Phe Thr Ser Cys Cys Pro Gly Trp
130 135 140

Val Ala Met Val Glu Lys Ser Asn Pro Glu Leu Ile Pro Tyr Leu Ser
145 150 155 160

Ser Cys Lys Ser Pro Gln Met Met Leu Gly Ala Val Ile Lys Asn Tyr
165 170 175

Tyr Ala Gln Gln Val Gly Val Gln Pro Ser Asp Ile Cys Asn Val Ser
180 185 190

Val Met Pro Cys Val Arg Lys Gln Gly Glu Ala Asp Arg Glu Trp Phe
195 200 205

Asn Thr Thr Gly Ala Gly Leu Ala Arg Asp Val Asp His Val Val Thr
210 215 220

Thr Ala Glu Val Gly Lys Ile Phe Leu Glu Arg Gly Ile Lys Leu Asn
225 230 235 240

Glu Leu Pro Glu Ser Asn Phe Asp Asn Pro Ile Gly Glu Gly Thr Gly
245 250 255

Gly Ala Leu Leu Phe Gly Thr Thr Gly Gly Val Met Glu Ala Ala Leu
260 265 270

Arg Thr Val Tyr Glu Val Val Thr Gln Lys Pro Met Gly Arg Val Asp
275 280 285

Phe Glu Glu Val Arg Gly Leu Glu Gly Ile Lys Glu Ala Glu Ile Thr
290 295 300

Leu Lys Pro Gly Asp Asp Ser Pro Phe Lys Ala Phe Ala Gly Ala Asp
305 310 315 320

Gly Gln Gly Ile Thr Leu Lys Ile Ala Val Ala Asn Gly Leu Gly Asn
325 330 335

Ala Lys Lys Leu Ile Lys Ser Leu Ser Glu Gly Lys Ala Lys Tyr Asp
340 345 350

Phe Ile Glu Val Met Ala Cys Pro Gly Gly Cys Ile Gly Gly Gly Gly
355 360 365

Gln Pro Arg Ser Thr Asp Lys Gln Ile Leu Gln Lys Arg Gln Gln Ala
370 375 380

Met Tyr Asn Leu Asp Glu Arg Ser Thr Ile Arg Arg Ser His Asp Asn
385 390 395 400

Pro Phe Ile Gln Ala Leu Tyr Asp Lys Phe Leu Gly Ala Pro Asn Ser
405 410 415

His Lys Ala His Asp Leu Leu His Thr His Tyr Val Ala Gly Gly Ile
420 425 430

Pro Glu Glu Lys
435

<210> 6
<211> 574
<212> PRT
<213> Clostridium saccharobutylicum

<400> 6

```
Met Ile Asn Ile Val Ile Asp Glu Lys Thr Ile Gln Val Gln Glu Asn
1               5                  10                  15

Thr Thr Val Ile Gln Ala Ala Leu Ala Asn Gly Ile Asp Ile Pro Ser
            20                  25                  30

Leu Cys Tyr Leu Asn Glu Cys Gly Asn Val Gly Lys Cys Gly Val Cys
        35                  40                  45

Ala Val Glu Ile Glu Gly Lys Asn Asn Leu Ala Leu Ala Cys Ile Thr
    50                  55                  60

Lys Val Glu Glu Gly Met Val Val Lys Thr Asn Ser Glu Lys Val Gln
65                  70                  75                  80

Glu Arg Val Lys Met Arg Val Ala Thr Leu Leu Asp Lys His Glu Phe
                85                  90                  95

Lys Cys Gly Pro Cys Pro Arg Arg Glu Asn Cys Glu Phe Leu Lys Leu
            100                 105                 110

Val Ile Lys Thr Lys Ala Lys Ala Asn Lys Pro Phe Val Val Glu Asp
        115                 120                 125

Lys Ser Gln Tyr Ile Asp Ile Arg Ser Lys Ser Ile Val Ile Asp Arg
    130                 135                 140

Thr Lys Cys Val Leu Cys Gly Arg Cys Glu Ala Ala Cys Lys Thr Lys
145                 150                 155                 160

Thr Gly Thr Gly Ala Ile Ser Ile Cys Lys Ser Glu Ser Gly Arg Ile
                165                 170                 175

Val Gln Ala Thr Gly Gly Lys Cys Phe Asp Asp Thr Asn Cys Leu Leu
            180                 185                 190

Cys Gly Gln Cys Val Ala Ala Cys Pro Val Gly Ala Leu Thr Glu Lys
        195                 200                 205

Thr His Val Asp Arg Val Lys Glu Ala Leu Glu Asp Pro Asn Lys His
    210                 215                 220

Val Ile Val Ala Met Ala Pro Ser Ile Arg Thr Ser Met Gly Glu Leu
225                 230                 235                 240

Phe Lys Leu Gly Tyr Gly Val Asp Val Thr Gly Lys Leu Tyr Ala Ser
                245                 250                 255

Met Arg Ala Leu Gly Phe Asp Lys Val Phe Asp Ile Asn Phe Gly Ala
            260                 265                 270

Asp Met Thr Ile Met Glu Glu Ala Thr Glu Phe Ile Glu Arg Val Lys
        275                 280                 285

Asn Asn Gly Pro Phe Pro Met Phe Thr Ser Cys Cys Pro Ala Trp Val
    290                 295                 300
```

```
Arg Gln Val Glu Asn Tyr Tyr Pro Glu Phe Leu Glu Asn Leu Ser Ser
305                 310                 315                 320

Ala Lys Ser Pro Gln Gln Ile Phe Gly Ala Ala Ser Lys Thr Tyr Tyr
                325                 330                 335

Pro Gln Ile Ser Gly Ile Ser Ala Lys Asp Val Phe Thr Val Thr Ile
            340                 345                 350

Met Pro Cys Thr Ala Lys Lys Phe Glu Ala Asp Arg Glu Glu Met Tyr
        355                 360                 365

Asn Glu Gly Ile Lys Asn Ile Asp Ala Val Leu Thr Thr Arg Glu Leu
    370                 375                 380

Ala Lys Met Ile Lys Asp Ala Lys Ile Asn Phe Ala Asn Leu Glu Asp
385                 390                 395                 400

Glu Gln Ala Asp Pro Ala Met Gly Glu Tyr Thr Gly Ala Gly Val Ile
                405                 410                 415

Phe Gly Ala Thr Gly Gly Val Met Glu Ala Ala Leu Arg Thr Ala Lys
            420                 425                 430

Asp Phe Val Glu Asp Lys Asp Leu Thr Asp Ile Glu Tyr Thr Gln Ile
        435                 440                 445

Arg Gly Leu Gln Gly Ile Lys Glu Ala Thr Val Glu Ile Gly Gly Glu
    450                 455                 460

Asn Tyr Asn Val Ala Val Ile Asn Gly Ala Ala Asn Leu Ala Glu Phe
465                 470                 475                 480

Met Asn Ser Gly Lys Ile Leu Glu Lys Asn Tyr His Phe Ile Glu Val
                485                 490                 495

Met Ala Cys Pro Gly Gly Cys Val Asn Gly Gly Gly Gln Pro His Val
            500                 505                 510

Ser Ala Lys Glu Arg Glu Lys Val Asp Val Arg Thr Val Arg Ala Ser
        515                 520                 525

Val Leu Tyr Asn Gln Asp Lys Asn Leu Glu Lys Arg Lys Ser His Lys
    530                 535                 540

Asn Thr Ala Leu Leu Asn Met Tyr Tyr Asp Tyr Met Gly Ala Pro Gly
545                 550                 555                 560

Gln Gly Lys Ala His Glu Leu Leu His Leu Lys Tyr Asn Lys
                565                 570
```

<210> 7
<211> 421
<212> PRT
<213> Desulfovibrio vulgaris

<400> 1

Met Ser Arg Ile Glu Met Glu Lys Ile Phe Tyr Glu Asp His Ala Pro
1 5 10 15

Asp Pro Lys Ala Asp Pro Asp Lys Leu Phe Phe Ile Gln Ile Asp Glu
20 25 30

Ser Lys Cys Ile Gly Cys Asp Ser Cys Gln Gln Tyr Cys Pro Thr Gly
35 40 45

Ala Ile Phe Gly Asp Thr Gly Asp Ala His Lys Ile Pro His Glu Glu
50 55 60

Leu Cys Ile Asn Cys Gly Gln Cys Leu Thr His Cys Pro Val Gly Ala
65 70 75 80

Ile Tyr Glu Ser Gln Ser Trp Val Thr Glu Ile Glu Lys Lys Ile Lys
85 90 95

Ala Lys Asp Val Lys Val Ile Ala Met Pro Ala Pro Ala Val Arg Tyr
100 105 110

Ala Leu Gly Asp Ala Phe Gly Leu Pro Val Gly Thr Val Thr Thr Gly
115 120 125

Lys Met Phe Ser Ala Leu Lys Glu Leu Gly Phe Asp His Cys Trp Asp
130 135 140

Asn Glu Phe Thr Ala Asp Val Thr Ile Trp Glu Glu Gly Thr Glu Phe
145 150 155 160

Val Gln Arg Leu Thr Lys Lys Leu Asp Lys Pro Leu Pro Gln Phe Thr
165 170 175

Ser Cys Cys Pro Gly Trp His Lys Tyr Val Glu Ser Leu Tyr Pro Glu
180 185 190

Leu Phe Pro His Met Ser Ser Cys Lys Ser Pro Ile Gly Met Leu Gly
195 200 205

Thr Leu Ala Lys Thr Tyr Gly Ala Asp Arg Met Lys Tyr Asp Arg Ala
210 215 220

Lys Val Tyr Thr Val Ser Ile Met Pro Cys Thr Ala Lys Lys Tyr Glu
225 230 235 240

Gly Met Arg Pro Gln Leu Trp Asp Ser Gly His Lys Asp Ile Asp Ala
245 250 255

Thr Ile Asp Thr Arg Glu Leu Ala Tyr Met Ile Lys Lys Ala Lys Ile
260 265 270

Asp Phe Thr Lys Leu Pro Asp Gly Lys Arg Asp Thr Leu Met Gly Glu
275 280 285

Ser Thr Gly Gly Ala Thr Leu Phe Gly Val Thr Gly Gly Val Met Glu

290 295 300

Ala Ala Leu Arg Tyr Ala Tyr Gln Ala Val Thr Gly Lys Lys Pro Glu
305 310 315 320

Ser Met Asp Phe Lys Gly Val Arg Gly Leu Gln Gly Val Lys Glu Ala
325 330 335

Thr Val Asn Val Gly Gly Val Asp Val Lys Val Ala Val Val His Gly
340 345 350

Ala Arg Arg Phe His Asp Val Cys Glu Leu Val Lys Ala Gly Lys Ala
355 360 365

Pro Trp His Phe Ile Glu Phe Met Ala Cys Pro Gly Gly Cys Val Cys
370 375 380

Gly Gly Gly Gln Pro Val Met Pro Gly Val Leu Glu Ala Ala Asp Arg
385 390 395 400

Arg Ser Thr Arg Met Tyr Ala Gly Leu Lys Lys Arg Leu Ala Met Ala
405 410 415

Ser Ala Ser Arg Ala
420

<210> 8
<211> 124
<212> PRT
<213> Desulfovibrio vulgaris

<400> 8

Met Gln Ile Val Asn Leu Thr Arg Arg Gly Phe Leu Lys Ala Ala Cys
1 5 10 15

Val Val Thr Gly Gly Ala Leu Ile Ser Ile Arg Met Thr Gly Lys Ala
20 25 30

Val Ala Ala Ala Lys Gln Leu Lys Asp Tyr Met Met Asp Arg Ile Asn
35 40 45

Gly Val Tyr Gly Ala Asp Ala Lys Phe Pro Val Arg Ala Ser Gln Asp
50 55 60

Asn Val Gln Val Gln Lys Leu Tyr Ala Asp Phe Leu Glu Lys Pro Met
65 70 75 80

Ser His Lys Ala Glu Gln Leu Leu His Thr His Trp Val Asp Arg Ser
85 90 95

Lys Ala Ile Glu Arg Met Lys Ala Gln Gly Ala Tyr Pro Asn Pro Arg
100 105 110

Ala Lys Glu Phe Glu Gly Asn Thr Tyr Pro Tyr Glu
115 120

<210> 9

<211> 606
<212> PRT
<213> Desulfovibrio vulgaris

<400> 9

Met Asn Ala Phe Ile Asn Gly Lys Glu Val Arg Cys Glu Pro Gly Arg
1 5 10 15

Thr Ile Leu Glu Ala Ala Arg Glu Asn Gly His Phe Ile Pro Thr Leu
20 25 30

Cys Glu Leu Ala Asp Ile Gly His Ala Pro Gly Thr Cys Arg Val Cys
35 40 45

Leu Val Glu Ile Trp Arg Asp Lys Glu Ala Gly Pro Gln Ile Val Thr
50 55 60

Ser Cys Thr Thr Pro Val Glu Glu Gly Met Arg Ile Phe Thr Arg Thr
65 70 75 80

Pro Glu Val Arg Arg Met Gln Arg Leu Gln Val Glu Leu Leu Leu Ala
85 90 95

Asp His Asp His Asp Cys Ala Ala Cys Ala Arg His Gly Asp Cys Glu
100 105 110

Leu Gln Asp Val Ala Gln Phe Val Gly Leu Thr Gly Thr Arg His His
115 120 125

Phe Pro Asp Tyr Ala Arg Ser Arg Thr Arg Asp Val Ser Ser Pro Ser
130 135 140

Val Val Arg Asp Met Gly Lys Cys Ile Arg Cys Leu Arg Cys Val Ala
145 150 155 160

Val Cys Arg Asn Val Gln Gly Val Asp Ala Leu Val Val Thr Gly Asn
165 170 175

Gly Ile Gly Thr Glu Ile Gly Leu Arg His Asn Arg Ser Gln Ser Ala
180 185 190

Ser Asp Cys Val Gly Cys Gly Gln Cys Thr Leu Val Cys Pro Val Gly
195 200 205

Ala Leu Ala Gly Arg Asp Asp Val Glu Arg Val Ile Asp Tyr Leu Tyr
210 215 220

Asp Pro Glu Ile Val Thr Val Phe Gln Phe Ala Pro Ala Val Arg Val
225 230 235 240

Gly Leu Gly Glu Glu Phe Gly Leu Pro Pro Gly Ser Ser Val Glu Gly
245 250 255

Gln Val Pro Thr Ala Leu Arg Leu Leu Gly Ala Asp Val Val Leu Asp
260 265 270

Thr Asn Phe Ala Ala Asp Leu Val Ile Met Glu Glu Gly Thr Glu Leu

Leu Gln Arg Leu Arg Gly Gly Ala Lys Leu Pro Leu Phe Thr Ser Cys
290 295 300

Cys Pro Gly Trp Val Asn Phe Ala Glu Lys His Leu Pro Asp Ile Leu
305 310 315 320

Pro His Val Ser Thr Thr Arg Ser Pro Gln Gln Cys Leu Gly Ala Leu
325 330 335

Ala Lys Thr Tyr Leu Ala Arg Thr Met Asn Val Ala Pro Glu Arg Met
340 345 350

Arg Val Val Ser Leu Met Pro Cys Thr Ala Lys Lys Glu Glu Ala Ala
355 360 365

Arg Pro Glu Phe Arg Arg Asp Gly Val Arg Asp Val Asp Ala Val Leu
370 375 380

Thr Thr Arg Glu Phe Ala Arg Leu Leu Arg Arg Glu Gly Ile Asp Leu
385 390 395 400

Ala Gly Leu Glu Pro Ser Pro Cys Asp Asp Pro Leu Met Gly Arg Ala
405 410 415

Thr Gly Ala Ala Val Ile Phe Gly Thr Thr Gly Gly Val Met Glu Ala
420 425 430

Ala Leu Arg Thr Val Tyr His Val Leu Asn Gly Lys Glu Leu Ala Pro
435 440 445

Val Glu Leu His Ala Leu Arg Gly Tyr Glu Asn Val Arg Glu Ala Val
450 455 460

Val Pro Leu Gly Glu Gly Asn Gly Ser Val Lys Val Ala Val Val His
465 470 475 480

Gly Leu Lys Ala Ala Arg Gln Met Val Glu Ala Val Leu Ala Gly Lys
485 490 495

Ala Asp His Val Phe Val Glu Val Met Ala Cys Pro Gly Gly Cys Met
500 505 510

Asp Gly Gly Gly Gln Pro Arg Ser Lys Arg Ala Tyr Asn Pro Asn Ala
515 520 525

Gln Ala Arg Arg Ala Ala Leu Phe Ser Leu Asp Ala Glu Asn Ala Leu
530 535 540

Arg Gln Ser His Asn Asn Pro Leu Ile Gly Lys Val Tyr Glu Ser Phe
545 550 555 560

Leu Gly Glu Pro Cys Ser Asn Leu Ser His Arg Leu Leu His Thr Arg
565 570 575

Tyr Gly Asp Arg Lys Ser Glu Val Ala Tyr Thr Met Arg Asp Ile Trp
580 585 590

His Glu Met Thr Leu Gly Arg Arg Val Arg Gly Asp Ser Asp
595 600 605

<210> 10
<211> 572
<212> PRT
<213> Clostridium perfringens

<400> 10

Met Asn Lys Ile Ile Ile Asn Asp Lys Thr Ile Glu Phe Asp Gly Asp
1 5 10 15

Lys Thr Ile Leu Asp Leu Ala Arg Glu Asn Gly Phe Asp Ile Pro Val
20 25 30

Leu Cys Glu Leu Lys Asn Cys Gly Asn Lys Gly Gln Cys Gly Val Cys
35 40 45

Leu Val Glu Gln Glu Gly Asn Asp Arg Leu Leu Arg Ser Cys Ala Ile
50 55 60

Lys Ala Lys Asp Gly Met Val Ile Lys Thr Asp Ser Glu Lys Val Leu
65 70 75 80

Glu Ala Arg Lys Glu Arg Val Ala Glu Leu Leu Asp Glu His Glu Phe
85 90 95

Lys Cys Gly Pro Cys Lys Arg Arg Glu Asn Cys Glu Phe Leu Lys Leu
100 105 110

Val Ile Lys Thr Lys Ala Arg Ala His Lys Pro Phe Val Val Ala Asp
115 120 125

Lys Ser Glu Tyr Val Asp Asp Arg Ser Lys Ser Ile Val Leu Asp Arg
130 135 140

Ser Lys Cys Val Lys Cys Gly Arg Cys Val Ala Ala Cys Arg Thr Arg
145 150 155 160

Thr Ala Thr Asn Ser Ile Lys Phe His Arg Ile Asp Gly Val Arg Leu
165 170 175

Val Gly Pro Glu Glu Leu Lys Cys Phe Asp Asp Thr Asn Cys Leu Leu
180 185 190

Cys Gly Gln Cys Ile Ala Ala Cys Pro Val Asp Ala Leu Ser Glu Lys
195 200 205

Ser His Ile Glu Arg Val Gln Glu Ala Leu Asn Asp Pro Glu Lys His
210 215 220

Val Ile Val Ala Met Ala Pro Ala Val Arg Thr Ser Met Gly Glu Leu
225 230 235 240

Phe Lys Met Gly Tyr Gly Gln Asp Val Thr Gly Lys Leu Tyr Thr Ala
245 250 255

Leu Arg Glu Leu Gly Phe Asp Lys Val Phe Asp Ile Asn Phe Gly Ala
260 265 270

Asp Met Thr Ile Met Glu Glu Ala Thr Glu Leu Ile Glu Arg Ile Lys
275 280 285

Asn Asn Gly Pro Phe Pro Met Leu Thr Ser Cys Cys Pro Ser Trp Val
290 295 300

Arg Glu Val Glu Asn Tyr Phe Pro Glu Leu Val Glu Asn Leu Ser Ser
305 310 315 320

Ala Lys Ser Pro Gln Gln Ile Phe Gly Ala Ala Ser Lys Thr Tyr Tyr
325 330 335

Pro Gln Val Ala Asp Ile Asp Pro Lys Lys Val Phe Thr Val Thr Val
340 345 350

Met Pro Cys Thr Ser Lys Lys Phe Glu Ala Asp Arg Pro Glu Met Glu
355 360 365

Asn Glu Gly Ile Arg Asn Ile Asp Ala Val Ile Thr Thr Arg Glu Leu
370 375 380

Ala Arg Met Ile Lys Ala Ala Lys Ile Asp Phe Ala Lys Leu Glu Asp
385 390 395 400

Gly Glu Val Asp Pro Ala Met Gly Glu Tyr Thr Gly Ala Gly Val Ile
405 410 415

Phe Gly Ala Thr Gly Gly Val Met Glu Ala Ala Leu Arg Thr Ala Lys
420 425 430

Asp Phe Met Glu Asn Asp Asn Leu Asp Asn Val Asp Tyr Glu Ala Val
435 440 445

Arg Gly Leu Ala Gly Ile Lys Glu Ala Glu Val Glu Ile Ala Gly Asn
450 455 460

Glu Tyr Lys Leu Ala Val Val Ser Gly Ala Ala Asn Val Phe Glu Leu
465 470 475 480

Val Lys Ser Gly Lys Ile Asn Asp Tyr His Phe Ile Glu Val Met Ala
485 490 495

Cys Pro Gly Gly Cys Val Asn Gly Gly Gly Gln Pro His Ile Ser Ala
500 505 510

Glu Asp Ser Asp Lys Met Asp Ile Arg Glu Val Arg Ala Ser Val Leu
515 520 525

Tyr Asn Gln Asp Lys Asn Leu Glu Lys Arg Lys Ser His Gln Asn Ser
530 535 540

Ala Leu Leu Lys Met Tyr Glu Ser Tyr Met Gly Lys Pro Gly His Gly
545 550 555 560

Arg Ala His Glu Leu Leu His Met Lys Tyr Lys Lys
565 570

<210> 11
<211> 572
<212> PRT
<213> Clostridium perfringens

<400> 11

Met Asn Lys Ile Ile Ile Asn Asp Lys Thr Ile Glu Phe Asp Gly Asp
1 5 10 15

Lys Thr Ile Leu Asp Leu Ala Arg Glu Asn Gly Phe Asp Ile Pro Val
20 25 30

Leu Cys Glu Leu Lys Asn Cys Gly Asn Lys Gly Gln Cys Gly Val Cys
35 40 45

Leu Val Glu Gln Glu Gly Asn Asp Arg Leu Leu Arg Ser Cys Ala Ile
50 55 60

Lys Ala Lys Asp Gly Met Val Ile Lys Thr Asp Ser Glu Lys Val Leu
65 70 75 80

Glu Ala Arg Lys Glu Arg Val Ala Glu Leu Leu Asp Glu His Glu Phe
85 90 95

Lys Cys Gly Pro Cys Lys Arg Arg Glu Asn Cys Glu Phe Leu Lys Leu
100 105 110

Val Ile Lys Thr Lys Ala Arg Ala His Lys Pro Phe Val Val Ala Asp
115 120 125

Lys Ser Glu Tyr Val Asp Asp Arg Ser Lys Ser Ile Val Leu Asp Arg
130 135 140

Ser Lys Cys Val Lys Cys Gly Arg Cys Val Ala Ala Cys Arg Thr Arg
145 150 155 160

Thr Ala Thr Asn Ser Ile Lys Phe His Arg Ile Asp Gly Val Arg Leu
165 170 175

Val Gly Pro Glu Glu Leu Lys Cys Phe Asp Asp Thr Asn Cys Leu Leu
180 185 190

Cys Gly Gln Cys Ile Ala Ala Cys Pro Val Asp Ala Leu Ser Glu Lys
195 200 205

Ser His Ile Glu Arg Val Gln Asp Ala Leu Asn Asp Pro Glu Lys His
210 215 220

Val Ile Val Ala Met Ala Pro Ala Val Arg Thr Ser Met Gly Glu Leu
225 230 235 240

Phe Lys Met Gly Tyr Gly Gln Asp Val Thr Gly Lys Leu Tyr Thr Ala
245 250 255

Leu Arg Glu Leu Gly Phe Asp Lys Val Phe Asp Ile Asn Phe Gly Ala
260 265 270

Asp Met Thr Ile Met Glu Glu Ala Thr Glu Leu Ile Glu Arg Ile Lys
275 280 285

Asn Asn Gly Pro Phe Pro Met Leu Thr Ser Cys Cys Pro Ser Trp Val
290 295 300

Arg Glu Val Glu Asn Tyr Phe Pro Glu Leu Val Glu Asn Leu Ser Ser
305 310 315 320

Ala Lys Ser Pro Gln Gln Ile Phe Gly Ala Ala Ser Lys Thr Tyr Tyr
325 330 335

Pro Gln Val Ala Asp Ile Asp Pro Lys Lys Val Phe Thr Val Thr Val
340 345 350

Met Pro Cys Thr Ser Lys Lys Phe Glu Ala Asp Arg Pro Glu Met Glu
355 360 365

Asn Glu Gly Ile Arg Asn Ile Asp Ala Val Ile Thr Thr Arg Glu Leu
370 375 380

Ala Arg Met Ile Lys Ala Ala Lys Ile Asp Phe Ala Lys Leu Glu Asp
385 390 395 400

Gly Glu Val Asp Pro Ala Met Gly Glu Tyr Thr Gly Ala Gly Val Ile
405 410 415

Phe Gly Ala Thr Gly Gly Val Met Glu Ala Ala Leu Arg Thr Ala Lys
420 425 430

Asp Phe Met Glu Asn Asp Asn Leu Asp Asn Val Asp Tyr Glu Ala Val
435 440 445

Arg Gly Leu Ala Gly Ile Lys Glu Ala Glu Val Glu Ile Ala Gly Asn
450 455 460

Glu Tyr Lys Leu Ala Val Val Ser Gly Ala Ala Asn Val Phe Glu Leu
465 470 475 480

Val Lys Ser Gly Lys Ile Asn Asp Tyr His Phe Ile Glu Val Met Ala
485 490 495

Cys Pro Gly Gly Cys Val Asn Gly Gly Gly Gln Pro His Ile Ser Ala
500 505 510

Glu Asp Ser Asp Lys Ile Asp Ile Arg Glu Val Arg Ala Ser Val Leu
515 520 525

Tyr Asn Gln Asp Lys Asn Leu Glu Lys Arg Lys Ser His Gln Asn Ser
530 535 540

Ala Leu Leu Lys Met Tyr Glu Asn Tyr Met Gly Lys Pro Gly His Gly
545 550 555 560

Arg Ala His Glu Leu Leu His Met Lys Tyr Lys Lys
565 570

<210> 12
<211> 484
<212> PRT
<213> Megasphaera elsdenii

<400> 12

Met Pro Glu Phe His Ser Arg Phe Glu Lys Ile Asp Arg Arg Val Pro
1 5 10 15

Ile Asp Glu His Asn Cys Ala Val Gln Phe Asp Val Thr Lys Cys Lys
20 25 30

Asn Cys Thr Leu Cys Arg Arg Ala Cys Ala Asp Thr Gln Thr Val Leu
35 40 45

Asp Tyr Tyr Ser Leu Ser Ser Thr Gly Asp Met Pro Ile Cys Val His
50 55 60

Cys Gly Gln Cys Ser Ser Ala Cys Pro Phe Gly Ala Ile Val Glu Val
65 70 75 80

Asn Asp Val Asp Lys Val Lys Ala Ala Leu Lys Asp Pro Glu Lys Ile
85 90 95

Val Ile Phe Gln Thr Ala Pro Ala Val Arg Val Gly Leu Gly Glu Ala
100 105 110

Phe Gly Met Asp Pro Gly Thr Phe Val Glu Gly Lys Met Val Ala Ala
115 120 125

Leu Arg Thr Leu Gly Ala Asp Tyr Val Phe Asp Thr Asp Phe Gly Ala
130 135 140

Asp Leu Thr Ile Met Glu Glu Ala Thr Glu Leu Leu His Arg Leu Gln
145 150 155 160

Ser Glu Glu Ile Pro Ile Pro Gln Phe Thr Ser Cys Cys Pro Ala Trp
165 170 175

Val Glu Phe Ala Glu Thr Phe Tyr Pro Asp Leu Leu Gln His Leu Ser
180 185 190

Ser Thr Lys Ser Pro Ile Ser Ile Leu Ser Pro Val Ile Lys Thr Tyr
195 200 205

Phe Ala Gln Gln Lys Asn Ile Asp Pro Lys Lys Ile Val Asn Val Cys
210 215 220

Val Thr Pro Cys Thr Ala Lys Lys Ala Glu Ile Arg Arg Pro Glu Leu
225 230 235 240

Ser Ala Ser Gly Leu Phe Trp Asp Glu Pro Glu Ile Arg Asp Thr Asp
245 250 255

Ile Cys Ile Thr Thr Arg Glu Leu Ala Gln Trp Ile Gln Asp Glu Asn
260 265 270

Ile Asp Phe Ala Ser Leu Glu Asp Ser Lys Phe Asp Lys Ala Phe Gly
275 280 285

Glu Ala Ser Gly Gly Gly Arg Ile Phe Gly Asn Ser Gly Gly Val Met
290 295 300

Glu Ala Ala Ile Arg Thr Ala Tyr His Met Phe Thr Gly Arg Pro Ala
305 310 315 320

Pro Lys Asp Phe Ile Pro Phe Glu Pro Val Arg Gly Leu Gln Gly Val
325 330 335

Lys Lys Ala Thr Val Ile Phe Gly His Phe Val Leu His Val Ala Ala
340 345 350

Ile Ser Gly Leu Gly Asn Ala Arg Ala Phe Ile Asp Asp Leu Ile Lys
355 360 365

Asn Asp Ala Phe Glu Asp Tyr Ser Phe Ile Glu Val Met Ala Cys Pro
370 375 380

Gly Gly Cys Ile Gly Gly Gly Gly Gln Pro Lys Val Lys Leu Pro Gln
385 390 395 400

Val Lys Lys Val Gln Glu Ala Arg Thr Ala Ser Ile Tyr Lys Ser Asp
405 410 415

Glu Glu Thr Asp Ile Lys Ala Ser Trp Gln Asn Pro Glu Ile Glu Thr
420 425 430

Leu Tyr Glu Ala Phe Leu Asp Glu Pro Leu Ser Glu Met Ala Glu Phe
435 440 445

Thr Leu His Thr Tyr Phe Ser Asp Lys Ser Asp Gln Leu Gly Arg Met
450 455 460

Lys Asn Leu Thr Pro Gln Thr Asn Pro Met Ser Pro Lys Tyr Lys Pro
465 470 475 480

Pro Thr Glu Glu

<210> 13
<211> 421
<212> PRT
<213> Desulfovibrio desulfuricans strain

<400> 13

Met Asn Leu Val Glu Met Glu Lys Ile Gln Tyr Val Asp Gln Ser Pro
1 5 10 15

Asp Pro Arg Ala Asn Pro Asp Glu Leu Phe Phe Ile Gln Ile Asp Pro
20 25 30

Glu Lys Cys Ile Gly Cys Asp Thr Cys Gln Glu Tyr Cys Pro Thr Gly
35 40 45

Ala Ile Phe Gly Asp Thr Gly Ser Ala His Ser Ile Pro His Glu Glu
50 55 60

Ile Cys Ile Asn Cys Gly Gln Cys Leu Thr His Cys Pro Val Gly Ala
65 70 75 80

Ile Tyr Glu Val Gln Ser Trp Val Arg Glu Leu Ser Glu Lys Ile Lys
85 90 95

Asp Pro Glu Ile Lys Val Ile Ala Met Pro Ala Pro Ala Val Arg Tyr
100 105 110

Gly Leu Gly Glu Cys Phe Gly Met Pro Val Gly Thr Val Thr Thr Gly
115 120 125

Lys Met Leu Thr Ala Leu Gln Met Leu Gly Phe Asp His Val Trp Asp
130 135 140

Asn Glu Phe Thr Ala Asp Val Thr Ile Trp Glu Glu Gly Thr Glu Phe
145 150 155 160

Val Asn Arg Leu Thr Gly Gln Ile Asp Lys Pro Leu Pro Gln Phe Thr
165 170 175

Ser Cys Cys Pro Gly Trp His Lys Tyr Val Glu Ser Phe Tyr Pro Glu
180 185 190

Leu Phe Pro His Leu Ser Ser Cys Lys Ser Pro Ile Gly Met Met Gly
195 200 205

Ala Leu Ala Lys Thr Tyr Gly Pro Asp Val Met Lys Tyr Asp Arg Ser
210 215 220

Lys Val Tyr Thr Val Ser Ile Met Pro Cys Thr Ala Lys Lys Tyr Glu
225 230 235 240

Gly Met Arg Ala Asp Leu Trp Ser Ser Gly Tyr Lys Asp Ile Asp Ala
245 250 255

Thr Ile Asp Thr Arg Glu Leu Ala Tyr Met Ile Lys Lys Ala Gly Ile
260 265 270

Asp Phe Ala Ala Leu Pro Asp Gly Lys Arg Asp Thr Leu Met Gly Asp
275 280 285

Ser Thr Gly Gly Ala Thr Ile Phe Gly Val Ser Gly Gly Val Met Glu
290 295 300

Ala Ala Leu Arg Tyr Ala Tyr Glu Ala Val Thr Gly Lys Lys Pro Ser

315 320

Ser Trp Asp Phe Thr Met Val Arg Gly Leu Asn Gly Ile Lys Glu Gly
325 330 335

Thr Val Thr Ile Gly Asp Ala Lys Ile Asn Val Ala Val Val His Gly
340 345 350

Ala Lys Arg Phe Ala Glu Val Cys Glu Val Ile Lys Thr Gly Lys Ser
355 360 365

Pro Trp His Phe Ile Glu Phe Met Ala Cys Pro Gly Gly Cys Val Cys
370 375 380

Gly Gly Gly Gln Pro Val Met Pro Gly Val Leu Glu Ala Met Asp Arg
385 390 395 400

Lys Val Ser Arg Thr Phe Ala Gly Leu Lys Glu Arg Leu Asn Arg Met
405 410 415

Ser Ser Ser Lys Ala
420

<210> 14
<211> 585
<212> PRT
<213> Desulfovibrio fructosovorans

<400> 14

Met Ser Met Leu Thr Ile Thr Ile Asp Gly Lys Thr Thr Ser Val Pro
1 5 10 15

Glu Gly Ser Thr Ile Leu Asp Ala Ala Lys Thr Leu Asp Ile Asp Ile
20 25 30

Pro Thr Leu Cys Tyr Leu Asn Leu Glu Ala Leu Ser Ile Asn Asn Lys
35 40 45

Ala Ala Ser Cys Arg Val Cys Val Val Glu Val Glu Gly Arg Arg Asn
50 55 60

Leu Ala Pro Ser Cys Ala Thr Pro Val Thr Asp Asn Met Val Val Lys
65 70 75 80

Thr Asn Ser Leu Arg Val Leu Asn Ala Arg Arg Thr Val Leu Glu Leu
85 90 95

Leu Leu Ser Asp His Pro Lys Asp Cys Leu Val Cys Ala Lys Ser Gly
100 105 110

Glu Cys Glu Leu Gln Thr Leu Ala Glu Arg Phe Gly Ile Arg Glu Ser
115 120 125

Pro Tyr Asp Gly Gly Glu Met Ser His Tyr Arg Lys Asp Ile Ser Ala
130 135 140

Ser Ile Ile Arg Asp Met Asp Lys Cys Ile Met Cys Arg Arg Cys Glu

145 150 155 160

Thr Met Cys Asn Thr Val Gln Thr Cys Gly Val Leu Ser Gly Val Asn
165 170 175

Arg Gly Phe Thr Ala Val Val Ala Pro Ala Phe Glu Met Asn Leu Ala
180 185 190

Asp Thr Val Cys Thr Asn Cys Gly Gln Cys Val Ala Val Cys Pro Thr
195 200 205

Gly Ala Leu Val Glu His Glu Tyr Ile Trp Glu Val Val Glu Ala Leu
210 215 220

Ala Asn Pro Asp Lys Val Val Ile Val Gln Thr Ala Pro Ala Val Arg
225 230 235 240

Ala Ala Leu Gly Glu Asp Leu Gly Val Ala Pro Gly Thr Ser Val Thr
245 250 255

Gly Lys Met Ala Ala Ala Leu Arg Arg Leu Gly Phe Asp His Val Phe
260 265 270

Asp Thr Asp Phe Ala Ala Asp Leu Thr Ile Met Glu Glu Gly Ser Glu
275 280 285

Phe Leu Asp Arg Leu Gly Lys His Leu Ala Gly Asp Thr Asn Val Lys
290 295 300

Leu Pro Ile Leu Thr Ser Cys Cys Pro Gly Trp Val Lys Phe Phe Glu
305 310 315 320

His Gln Phe Pro Asp Met Leu Asp Val Pro Ser Thr Ala Lys Ser Pro
325 330 335

Gln Gln Met Phe Gly Ala Ile Ala Lys Thr Tyr Tyr Ala Asp Leu Leu
340 345 350

Gly Ile Pro Arg Glu Lys Leu Val Val Val Ser Val Met Pro Cys Leu
355 360 365

Ala Lys Lys Tyr Glu Cys Ala Arg Pro Glu Phe Ser Val Asn Gly Asn
370 375 380

Pro Asp Val Asp Ile Val Ile Thr Thr Arg Glu Leu Ala Lys Leu Val
385 390 395 400

Lys Arg Met Asn Ile Asp Phe Ala Gly Leu Pro Asp Glu Asp Phe Asp
405 410 415

Ala Pro Leu Gly Ala Ser Thr Gly Ala Ala Pro Ile Phe Gly Val Thr
420 425 430

Gly Gly Val Ile Glu Ala Ala Leu Arg Thr Ala Tyr Glu Leu Ala Thr
435 440 445

Gly Glu Thr Leu Lys Lys Val Asp Phe Glu Asp Val Arg Gly Met Asp
450 455 460

Gly Val Lys Lys Ala Lys Val Lys Val Gly Asp Asn Glu Leu Val Ile
465 470 475 480

Gly Val Ala His Gly Leu Gly Asn Ala Arg Glu Leu Leu Lys Pro Cys
485 490 495

Gly Ala Gly Glu Thr Phe His Ala Ile Glu Val Met Ala Cys Pro Gly
500 505 510

Gly Cys Ile Gly Gly Gly Gly Gln Pro Tyr His His Gly Asp Val Glu
515 520 525

Leu Leu Lys Lys Arg Thr Gln Val Leu Tyr Ala Glu Asp Ala Gly Lys
530 535 540

Pro Leu Arg Lys Ser His Glu Asn Pro Tyr Ile Ile Glu Leu Tyr Glu
545 550 555 560

Lys Phe Leu Gly Lys Pro Leu Ser Glu Arg Ser His Gln Leu Leu His
565 570 575

Thr His Tyr Phe Lys Arg Gln Arg Leu
580 585

<210> 15
<211> 421
<212> PRT
<213> Desulfovibrio fructosovorans

<400> 15

Met Ser Arg Ile Glu Met Ala Lys Ile Phe Tyr Glu Gln Thr Val Pro
1 5 10 15

Pro Pro Gly Thr Asn Leu Asp Gln Ala Tyr Ile Val Gln Val Asp Glu
20 25 30

Thr Lys Cys Ile Gly Cys Asp Thr Cys Met Gly Tyr Cys Pro Thr Gly
35 40 45

Ala Ile Thr Gly Glu Ser Gly Glu Pro His Lys Val Val Asp Pro Ala
50 55 60

Ala Cys Ile Asn Cys Gly Gln Cys Leu Thr His Cys Pro Val Ala Ala
65 70 75 80

Ile Tyr Glu Thr Val Ser Phe Val Pro Glu Ile Glu Ala Lys Leu Lys
85 90 95

Asp Lys Asn Val Lys Val Ile Ala Met Pro Ala Pro Ala Val Arg Tyr
100 105 110

Ala Leu Gly Asp Pro Phe Gly Met Pro Leu Gly Ala Val Thr Thr Glu
115 120 125

PCHTs Met Leu Thr Gly Leu Lys Gln Leu Gly Phe Asp Asn Val Trp Asp
130 135 140

Asn Glu Phe Thr Ala Asp Val Thr Ile Trp Glu Glu Gly Ser Glu Leu
145 150 155 160

Leu Ala Arg Ile Thr Lys Lys Leu Asp Lys Pro Leu Pro Gln Phe Thr
165 170 175

Ser Cys Cys Pro Gly Trp Gln Lys Tyr Ala Glu Thr Phe Tyr Pro Glu
180 185 190

Leu Leu Pro His Phe Ser Ser Cys Lys Ser Pro Ile Gly Met Met Gly
195 200 205

Pro Leu Ala Lys Thr Tyr Gly Ala Lys Glu Leu Gly Tyr Glu Pro Lys
210 215 220

Gln Ile Tyr Thr Val Ser Ile Met Pro Cys Thr Ala Lys Lys Phe Glu
225 230 235 240

Gly Met Arg Pro Glu Met Asp Ala Ser Gly Phe Arg Asp Ile Asp Ala
245 250 255

Thr Ile Asn Thr Arg Glu Leu Ala Tyr Met Met Lys Lys Ala Gly Ile
260 265 270

Asp Leu Pro Lys Ile Ala Asn Gly Lys Arg Asp Ala Val Met Gly Glu
275 280 285

Ser Thr Gly Gly Ala Thr Ile Phe Gly Val Ser Gly Gly Val Met Glu
290 295 300

Ala Ala Leu Arg Phe Ala Tyr Gln Ala Leu Thr Lys Lys Pro Pro Gln
305 310 315 320

Ser Trp Asp Phe Lys Ala Val Arg Gly Leu Asn Gly Ile Lys Glu Ala
325 330 335

Thr Ile Asn Ile Gly Gly Thr Asp Val Lys Val Ala Val Val Asn Gly
340 345 350

Gly Lys Asn Phe Ala Lys Val Cys Asp Glu Val Lys Ala Gly Lys Ser
355 360 365

Pro Tyr His Phe Ile Glu Phe Met Ala Cys Pro Gly Gly Cys Val Met
370 375 380

Gly Gly Gly Gln Pro Ile Met Pro Thr Val Leu Glu Ser Met Asn Arg
385 390 395 400

Thr Thr Thr Lys Phe Tyr Ala Ser Leu Lys Lys Arg Leu Ala Leu Tyr
405 410 415

Asp Ala Gln Lys Ala
420

<210> 16
<211> 608
<212> PRT
<213> Thermotoga maritima

<400> 16

Met Arg Arg Phe Phe Lys Asn Asn Leu Arg Asn Leu Ser Gln Asn Gly
1 5 10 15

Glu Thr Asn Ser Val Arg Arg Cys Phe Ala Leu Ala Asp Val Thr Val
20 25 30

Val Ile Asn Gly Arg Thr Leu Thr Val Pro Asp Asn Leu Thr Val Ile
35 40 45

Glu Ala Cys Glu Lys Ala Gly Ile Glu Ile Pro Ala Leu Cys His His
50 55 60

Pro Arg Leu Gly Glu Ser Ile Gly Ala Cys Arg Val Cys Val Val Glu
65 70 75 80

Val Glu Gly Ala Arg Asn Leu Gln Pro Ala Cys Val Thr Lys Val Arg
85 90 95

Asp Gly Met Val Ile Lys Thr Ser Ser Asp Arg Val Lys Thr Ala Arg
100 105 110

Lys Phe Asn Leu Ala Leu Leu Leu Ser Glu His Pro Asn Asp Cys Met
115 120 125

Thr Cys Glu Ala Asn Gly Arg Cys Glu Phe Gln Asp Leu Ile Tyr Lys
130 135 140

Tyr Asp Val Glu Pro Ile Phe Gly Tyr Gly Thr Lys Glu Gly Leu Val
145 150 155 160

Asp Arg Ser Ser Pro Ala Ile Val Arg Asp Leu Ser Lys Cys Ile Lys
165 170 175

Cys Gln Arg Cys Val Arg Ala Cys Ser Glu Leu Gln Gly Met His Ile
180 185 190

Tyr Ser Met Val Glu Arg Gly His Arg Thr Tyr Pro Gly Thr Pro Phe
195 200 205

Asp Met Pro Val Tyr Glu Thr Asp Cys Ile Gly Cys Gly Gln Cys Ala
210 215 220

Ala Phe Cys Pro Thr Gly Ala Ile Val Glu Asn Ser Ala Val Lys Val
225 230 235 240

Val Leu Glu Glu Leu Glu Lys Lys Glu Lys Ile Leu Val Val Gln Thr
245 250 255

Ala Pro Ser Val Arg Val Ala Ile Gly Glu Glu Phe Gly Tyr Ala Pro
260 265 270

Gly Thr Ile Ser Thr Gly Gln Met Val Ala Ala Leu Arg Arg Leu Gly
275 280 285

Phe Asp Tyr Val Phe Asp Thr Asn Phe Gly Ala Asp Leu Thr Ile Met
290 295 300

Glu Glu Gly Ser Glu Phe Leu Glu Arg Leu Glu Lys Gly Asp Leu Glu
305 310 315 320

Asp Leu Pro Met Phe Thr Ser Cys Cys Pro Gly Trp Val Asn Leu Val
325 330 335

Glu Lys Val Tyr Pro Glu Leu Arg Thr Arg Leu Ser Ser Ala Lys Ser
340 345 350

Pro Gln Gly Met Leu Ser Ala Met Val Lys Thr Tyr Phe Ala Glu Lys
355 360 365

Leu Gly Val Lys Pro Glu Asp Ile Phe His Val Ser Ile Met Pro Cys
370 375 380

Thr Ala Lys Lys Asp Glu Ala Leu Arg Lys Gln Leu Met Val Asn Gly
385 390 395 400

Val Pro Ala Val Asp Val Val Leu Thr Thr Arg Glu Leu Gly Lys Leu
405 410 415

Ile Arg Met Lys Lys Ile Pro Phe Ala Asn Leu Pro Glu Glu Glu Tyr
420 425 430

Asp Ala Pro Leu Gly Ile Ser Thr Gly Ala Ala Ala Leu Phe Gly Val
435 440 445

Thr Gly Gly Val Met Glu Ala Ala Leu Arg Thr Ala Tyr Glu Leu Lys
450 455 460

Thr Gly Lys Ala Leu Pro Lys Ile Val Phe Glu Glu Val Arg Gly Leu
465 470 475 480

Lys Gly Val Arg Glu Ala Glu Ile Asp Leu Asp Gly Lys Lys Ile Arg
485 490 495

Ile Ala Val Val His Gly Thr Ala Asn Val Arg Asn Leu Val Glu Lys
500 505 510

Ile Leu Arg Arg Glu Val Lys Tyr His Phe Val Glu Val Met Ala Cys
515 520 525

Pro Gly Gly Cys Ile Gly Gly Gly Gly Gln Pro Tyr Ser Arg Asp Pro
530 535 540

Glu Ile Leu Arg Lys Arg Ala Glu Ala Ile Tyr Thr Ile Asp Glu Arg
545 550 555 560

Met Thr Leu Arg Lys Ser His Glu Asn Pro Ala Ile Lys Lys Leu Tyr
565 570 575

Glu Glu Tyr Leu Glu His Pro Leu Ser His Lys Ala His Glu Leu Leu
580 585 590

His Thr Tyr Tyr Glu Asp Arg Ser Arg Lys Lys Arg Leu Ala Val Lys
595 600 605

<210> 17
<211> 645
<212> PRT
<213> Thermotoga maritima

<400> 17

Met Lys Ile Tyr Val Asp Gly Arg Glu Val Ile Ile Asn Asp Asn Glu
1 5 10 15

Arg Asn Leu Leu Glu Ala Leu Lys Asn Val Gly Ile Glu Ile Pro Asn
20 25 30

Leu Cys Tyr Leu Ser Glu Ala Ser Ile Tyr Gly Ala Cys Arg Met Cys
35 40 45

Leu Val Glu Ile Asn Gly Gln Ile Thr Thr Ser Cys Thr Leu Lys Pro
50 55 60

Tyr Glu Gly Met Lys Val Lys Thr Asn Thr Pro Glu Ile Tyr Glu Met
65 70 75 80

Arg Arg Asn Ile Leu Glu Leu Ile Leu Ala Thr His Asn Arg Asp Cys
85 90 95

Thr Thr Cys Asp Arg Asn Gly Ser Cys Lys Leu Gln Lys Tyr Ala Glu
100 105 110

Asp Phe Gly Ile Arg Lys Ile Arg Phe Glu Ala Leu Lys Lys Glu His
115 120 125

Val Arg Asp Glu Ser Ala Pro Val Val Arg Asp Thr Ser Lys Cys Ile
130 135 140

Leu Cys Gly Asp Cys Val Arg Val Cys Glu Glu Ile Gln Gly Val Gly
145 150 155 160

Val Ile Glu Phe Ala Lys Arg Gly Phe Glu Ser Val Val Thr Thr Ala
165 170 175

Phe Asp Thr Pro Leu Ile Glu Thr Glu Cys Val Leu Cys Gly Gln Cys
180 185 190

Val Ala Tyr Cys Pro Thr Gly Ala Leu Ser Ile Arg Asn Asp Ile Asp
195 200 205

Lys Leu Ile Glu Ala Leu Glu Ser Asp Lys Ile Val Ile Gly Met Ile
210 215 220

Ala Pro Ala Val Arg Ala Ala Ile Gln Glu Glu Phe Gly Ile Asp Glu
225 230 235 240

Asp Val Ala Met Ala Glu Lys Leu Val Ser Phe Leu Lys Thr Ile Gly
245 250 255

Phe Asp Lys Val Phe Asp Val Ser Phe Gly Ala Asp Leu Val Ala Tyr
260 265 270

Glu Glu Ala His Glu Phe Tyr Glu Arg Leu Lys Lys Gly Glu Arg Leu
275 280 285

Pro Gln Phe Thr Ser Cys Cys Pro Ala Trp Val Lys His Ala Glu His
290 295 300

Thr Tyr Pro Gln Tyr Leu Gln Asn Leu Ser Ser Val Lys Ser Pro Gln
305 310 315 320

Gln Ala Leu Gly Thr Val Ile Lys Lys Ile Tyr Ala Arg Lys Leu Gly
325 330 335

Val Pro Glu Glu Lys Ile Phe Leu Val Ser Phe Met Pro Cys Thr Ala
340 345 350

Lys Lys Phe Glu Ala Glu Arg Glu Glu His Glu Gly Ile Val Asp Ile
355 360 365

Val Leu Thr Thr Arg Glu Leu Ala Gln Leu Ile Lys Met Ser Arg Ile
370 375 380

Asp Ile Asn Arg Val Glu Pro Gln Pro Phe Asp Arg Pro Tyr Gly Val
385 390 395 400

Ser Ser Gln Ala Gly Leu Gly Phe Gly Lys Ala Gly Gly Val Phe Ser
405 410 415

Cys Val Leu Ser Val Leu Asn Glu Glu Ile Gly Ile Glu Lys Val Asp
420 425 430

Val Lys Ser Pro Glu Asp Gly Ile Arg Val Ala Glu Val Thr Leu Lys
435 440 445

Asp Gly Thr Ser Phe Lys Gly Ala Val Ile Tyr Gly Leu Gly Lys Val
450 455 460

Lys Lys Phe Leu Glu Glu Arg Lys Asp Val Glu Ile Ile Glu Val Met
465 470 475 480

Ala Cys Asn Tyr Gly Cys Val Gly Gly Gly Gly Gln Pro Tyr Pro Asn
485 490 495

Asp Ser Arg Ile Arg Glu His Arg Ala Lys Val Leu Arg Asp Thr Met
500 505 510

Gly Ile Lys Ser Leu Leu Thr Pro Val Glu Asn Leu Phe Leu Met Lys
515 520 525

Leu Tyr Glu Glu Asp Leu Lys Asp Glu His Thr Arg His Glu Ile Leu

530 535 540

His Thr Thr Tyr Arg Pro Arg Arg Arg Tyr Pro Glu Lys Asp Val Glu
545 550 555 560

Ile Leu Pro Val Pro Asn Gly Glu Lys Arg Thr Val Lys Val Cys Leu
565 570 575

Gly Thr Ser Cys Tyr Thr Lys Gly Ser Tyr Glu Ile Leu Lys Lys Leu
580 585 590

Val Asp Tyr Val Lys Glu Asn Asp Met Glu Gly Lys Ile Glu Val Leu
595 600 605

Gly Thr Phe Cys Val Glu Asn Cys Gly Ala Ser Pro Asn Val Ile Val
610 615 620

Asp Asp Lys Ile Ile Gly Gly Ala Thr Phe Glu Lys Val Leu Glu Glu
625 630 635 640

Leu Ser Lys Asn Gly
645

<210> 18
<211> 1206
<212> PRT
<213> Nyctotherus ovalis

<400> 18

Met Ile Ser Arg Leu Ile Ala Lys Lys Ala Pro Leu Phe Leu Arg Thr
1 5 10 15

Phe Ala Thr Ser Glu Met Ile Ser Leu Lys Ile Asp Gly Lys Ile Ile
20 25 30

Ser Val Pro Lys Gly Ile Met Leu Ala Asp Ala Ile Lys Lys Ala Gly
35 40 45

Ala Asn Val Pro Thr Met Cys Tyr His Pro Asp Leu Pro Thr Ser Gly
50 55 60

Gly Ile Cys Arg Val Cys Leu Val Glu Ser Ala Lys Ser Pro Gly Tyr
65 70 75 80

Pro Ile Ile Ser Cys Arg Thr Pro Val Glu Glu Gly Met Glu Ile Val
85 90 95

Thr Gln Gly Ser Lys Met Lys Glu Tyr Arg Gln Ala Asn Leu Ala Leu
100 105 110

Met Leu Ser Arg His Pro Asn Ala Cys Leu Ser Cys Thr Ser Asn Thr
115 120 125

Asn Cys Lys Thr Gln Glu Leu Ser Ala Asn Met Asn Ile Gly Gln Cys
130 135 140

Gly Phe Ala Asn Ala Thr Pro Pro Lys Asn Asp Asp Ser Tyr Asp Met

145 150 155 160

Thr Thr Ala Ile Glu Arg Asp Asn Asp Lys Cys Ile Asn Cys Asp Ile
165 170 175

Cys Val His Thr Cys Ser Leu Gln Gly Leu Asn Ala Leu Gly Phe Tyr
180 185 190

Asn Glu Glu Gly His Ala Val Lys Ser Met Gly Thr Leu Asp Val Ser
195 200 205

Glu Cys Ile Gln Cys Gly Gln Cys Ile Asn Arg Cys Pro Thr Gly Ala
210 215 220

Ile Thr Glu Lys Ser Glu Ile Arg Pro Val Leu Asp Ala Ile Asn Ile
225 230 235 240

Gln Gln Arg Leu Val Phe Gln Met Ala Pro Ser Ile Arg Val Ala Val
245 250 255

Ala Glu Glu Phe Gly Ile Lys Pro Gly Glu Lys Ile Leu Lys Asn Glu
260 265 270

Ile Ala Thr Ala Leu Arg Lys Leu Gly Ser Asn Val Phe Val Leu Asp
275 280 285

Thr Asn Phe Ser Ala Asp Leu Thr Ile Ile Glu Glu Gly His Glu Leu
290 295 300

Ile Glu Arg Leu Tyr Arg Asn Val Thr Gly Lys Lys Leu Leu Gly Gly
305 310 315 320

Asp His Met Pro Ile Asp Leu Pro Met Leu Thr Ser Cys Cys Pro Gly
325 330 335

Trp Ile Met Phe Ile Glu Lys Asn Tyr Pro Asp Leu Leu Asn Asn Leu
340 345 350

Ser Thr Cys Lys Ser Pro Gln Gly Met Leu Gly Ala Leu Ile Lys Gly
355 360 365

Tyr Trp Ala Lys Asn Ile Lys Lys Met Asp Pro Lys Asp Ile Val Ser
370 375 380

Val Ser Ile Met Pro Cys Thr Ala Lys Lys Ala Glu Lys Glu Arg Pro
385 390 395 400

Gln Leu Arg Gly Asp Glu Gly Tyr Lys Asp Val Asp Tyr Ile Leu Thr
405 410 415

Thr Arg Glu Leu Ala Lys Met Leu Lys Gln Ser Asn Ile Asp Leu Ala
420 425 430

Lys Met Glu Pro Thr Pro Phe Asp Lys Val Met Ser Glu Gly Thr Gly
435 440 445

Ala Ala Val Ile Phe Gly Val Thr Gly Gly Val Met Glu Ala Ala Leu
450 455 460

Arg Thr Ala Asn Glu Val Ile Thr Gly Arg Glu Val Pro Phe Lys Asn
465 470 475 480

Leu Asn Ile Glu Ala Val Arg Gly Met Glu Gly Ile Arg Glu Ala Gly
485 490 495

Ile Lys Leu Glu Asn Val Leu Asp Lys Tyr Lys Ala Phe Glu Gly Val
500 505 510

Thr Val Lys Val Ala Ile Ala His Gly Pro Asn Asn Ala Arg Lys Val
515 520 525

Met Asp Ile Ile Lys Gln Ala Lys Glu Ser Gly Lys Pro Ala Pro Trp
530 535 540

His Phe Val Glu Val Met Ala Cys Pro Gly Gly Cys Ile Gly Gly Gly
545 550 555 560

Gly Gln Pro Lys Pro Thr Asn Leu Glu Ile Arg Gln Ala Arg Thr Gln
565 570 575

Leu Thr Phe Lys Glu Asp Met Asp Leu Pro Leu Arg Lys Ser His Asp
580 585 590

Asn Pro Glu Ile Lys Ala Ile Tyr Glu Asn Tyr Leu Lys Glu Pro Leu
595 600 605

Gly His Asn Ser His His Tyr Leu His Thr Thr Tyr Ser Ser Gln Lys
610 615 620

Val Arg Asp Met Asn Leu Tyr Asn Ala Asn Glu Ala Ala Gly Leu Asp
625 630 635 640

Glu Ile Leu Ala Lys Tyr Pro Lys Glu Lys Glu Tyr Leu Met Pro Ile
645 650 655

Ile Ile Glu Glu His Asp Lys Lys Gly Tyr Ile Ser Asp Pro Ser Ile
660 665 670

Val Lys Ile Ser Glu His Leu Gly Met Tyr Pro Ala Gln Ile Glu Ser
675 680 685

Ile Leu Ser Ser Tyr His Tyr Phe Pro Arg Glu His Thr Ile Ala Ile
690 695 700

Leu Met Ser Ile Cys Val His Cys His Asn Cys Met Met Lys Gly Gln
705 710 715 720

Gly Arg Leu Leu Lys Thr Ile Gln Glu Thr Tyr Asp Ile His Glu Thr
725 730 735

His Gly Gly Val Ala Lys Asp Gly Ser Phe Thr Leu His Thr Leu Asn
740 745 750

```
Trp Leu Gly Tyr Cys Val Asn Asp Ala Pro Ala Met Met Ile Lys Arg
        755                 760                 765

Lys Gly Thr Asn Tyr Val Glu Thr Phe Thr Gly Leu Leu Gly Asp Asn
    770                 775                 780

Ile Asp Gln Arg Leu Lys Ser Leu Lys Asn Leu Lys Lys Glu Leu Pro
785                 790                 795                 800

Lys Trp Pro Lys Asn Asn Ile Arg Glu Met Lys Ser Gln Arg Asn Gly
                805                 810                 815

Asn Ser Tyr Ser Cys Met Asn Thr Gln Ala Pro Ile Ala Glu Ala Thr
            820                 825                 830

Lys Lys Ala Val Ser Met Gly Pro Glu Lys Val Ile Glu Glu Val Phe
        835                 840                 845

Lys Ser Asn Leu Val Gly Arg Gly Gly Ala Gly Phe Arg Thr Gly Lys
    850                 855                 860

Lys Trp Glu Ser Ala Tyr Lys Thr Pro Ala Ser Asp Lys Tyr Val Val
865                 870                 875                 880

Cys Asn Ala Asp Glu Gly Leu Pro Ser Thr Tyr Lys Asp Trp Cys Leu
                885                 890                 895

Leu Asn Asn Glu Ala Lys Arg Lys Glu Val Phe Thr Gly Met Gly Ile
            900                 905                 910

Cys Ala Lys Thr Ile Gly Ala Lys Arg Cys Phe Met Tyr Leu Arg Tyr
        915                 920                 925

Glu Tyr Arg Asn Leu Val Pro Ala Leu Glu Gln Ser Ile Lys Asp Val
    930                 935                 940

Gln Ser Thr Cys Pro Glu Leu Ala Asp Leu Lys Tyr Glu Ile Arg Leu
945                 950                 955                 960

Gly Gly Gly Pro Tyr Val Ala Gly Glu Glu Asn Ala Gln Phe Glu Ser
                965                 970                 975

Ile Glu Gly Arg Ala Pro Leu Pro Arg Lys Asp Arg Pro Gly Asn Ile
            980                 985                 990

Phe Pro Thr Met Glu Gly Leu Phe  His Lys Pro Thr Val  Ile Asn Asn
        995                 1000                1005

Val Glu  Thr Phe Phe Ala Ile  Pro His Ile Ile Gln  Gln Gly Ser
    1010                1015                1020

Gln Ser  Phe Gly Glu Gly Lys  Met Pro Lys Leu Leu  Ser Val Thr
    1025                1030                1035

Gly Asp  Val Asp Glu Pro Ile  Leu Ile Glu Thr Asn  Leu Asn Asn
    1040                1045                1050
```

Tyr Ser Leu Asn His Leu Leu Gln Glu Ile Ser Ala Lys Asp Ile
1055 1060 1065

Val Ala Ala Glu Ile Gly Gly Cys Thr Glu Pro Ile Ile Phe Gly
1070 1075 1080

Ser Lys Phe Asp Thr Leu Phe Gly Phe Gly Arg Gly Thr Leu Asn
1085 1090 1095

Ala Val Gly Ser Val Val Leu Phe Asn Ser Ser Cys Asp Leu Gly
1100 1105 1110

Lys Ile Tyr Glu Asn Lys Leu Lys Phe Met Ala Glu Glu Ser Cys
1115 1120 1125

Lys Gln Cys Val Pro Cys Arg Asp Gly Ser Tyr Ile Phe His Arg
1130 1135 1140

Ala Phe Lys Glu Leu Arg Asp Thr Gly Lys Ser Ser Tyr Asn Met
1145 1150 1155

Arg Ala Leu Ala Val Ala Ser Glu Ser Ala Ala Arg Ser Ser Ile
1160 1165 1170

Cys Ala His Gly Lys Ala Leu Glu Ser Leu Phe Lys Ser Ala Cys
1175 1180 1185

Asp Phe Met Asn Lys Thr Lys Pro Ile Tyr Gln Pro His Ser Thr
1190 1195 1200

Tyr His Gln
1205

<210> 19
<211> 467
<212> PRT
<213> Spironucleus barkhanus

<400> 19

Met Lys Val Arg Gln Ser Pro Phe Lys Ile Asp Ile Thr Asn Gly Pro
1 5 10 15

Ile Asp Arg Asn Asp Ala Ile Gln Ile Asp Tyr Gln Lys Cys Ile Gly
20 25 30

Cys Gln Met Cys Ala Lys Thr Cys Thr Asp Ser Gln Asn Phe Asn Ile
35 40 45

Phe Lys Ile Ser Ala Pro Lys Thr Lys Pro Phe Val Asn Ala Tyr Gly
50 55 60

Ser Val Ala Glu Gly Thr Glu Arg Asn Ala Leu Ala Gly Thr Asp Cys
65 70 75 80

Thr Gly Cys Gly Ala Cys Val Arg Ala Cys Pro Val Glu Ala Leu Met
85 90 95

```
Pro Ala Phe Asn Ile Arg Pro Val Leu Glu Pro Ile Ser Glu Lys Lys
            100                 105                 110

Lys Val Thr Ile Ala Val Ile Ala Pro Ser Thr Arg Val Gly Leu Ala
        115                 120                 125

Glu Gly Met Gly Met Gly Val Gly Val Thr Ala Glu Arg Gln Met Val
    130                 135                 140

Tyr Glu Leu Lys Gln Met Gly Phe Asp Tyr Val Phe Asp Asn Met Trp
145                 150                 155                 160

Gly Ala Asp Ala Pro Thr Thr Glu Asp Ala Lys Glu Ile Leu Lys Ala
                165                 170                 175

Lys Ala Ala Gly Lys Thr Ala Phe Thr Ser Cys Cys Pro Ala Trp Val
            180                 185                 190

Lys Leu Val Glu Thr Thr Tyr Pro Glu Leu Leu Pro Asn Ile Ser Ser
        195                 200                 205

Ala Arg Ser Pro His Gly Ile Ile Cys Ser Val Ile Lys Lys Tyr Phe
    210                 215                 220

Ala Lys Asp Ile Gly Lys Lys Ala Asp Glu Leu Tyr Val Val Gly Val
225                 230                 235                 240

Met Pro Cys Thr Ala Lys Lys Asn Glu Ala Ala Arg Lys Glu Leu Thr
                245                 250                 255

Thr Asp Gly Ser Pro Asp Cys Asp Ile Ser Ile Thr Thr Arg Glu Leu
            260                 265                 270

Met Ala Tyr Leu Lys Glu Lys Lys Val Thr Phe Ser Ala Ala Arg Glu
        275                 280                 285

Ile Glu Leu Lys Asp Asn Val Gln Ala Gln Tyr Asp Ala Pro Phe Asn
    290                 295                 300

Thr Phe Ser Gly Ser Ala Tyr Ile Tyr Gly Lys Thr Ala Gly Val Thr
305                 310                 315                 320

Glu Ala Val Val Arg Tyr Val Cys Ala Ile Lys Lys Val Pro Phe Ser
                325                 330                 335

Val Gly Met Ile Thr Lys Glu Leu Ile Trp Glu Asn Lys Leu His Ser
            340                 345                 350

Ser Ser Leu Thr Leu Leu Thr Phe Ser Ala Ala Gly Glu Asp Tyr Arg
        355                 360                 365

Ile Cys Val Ser Tyr Gly Gly Leu Ala Ala His Lys Ala Val Glu Leu
    370                 375                 380

Tyr Lys Ser Gly Glu Leu Lys Val Asp Ala Val Glu Val Met Val Cys
```

385 390 395 400

Pro Gly Gly Cys Val Gly Gly Gly Gly Gln Pro Lys Gln Pro Lys Lys
405 410 415

Asp Met Ile Leu Lys Arg His Glu Gly Leu Asp Lys His Asp Lys Glu
420 425 430

Ala Pro Tyr Ser Asn Cys Thr Glu Asn Pro Thr Leu Asn Glu Phe Tyr
435 440 445

Glu Arg Ile Gly Thr Asp Val His His Val Met His Thr Thr Tyr Ser
450 455 460

Ala Tyr Lys
465

<210> 20
<211> 468
<212> PRT
<213> Trichomonas vaginalis

<400> 20

Met Leu Ala Ser Ser Ala Thr Ala Met Lys Gly Phe Ala Asn Ser Leu
1 5 10 15

Arg Met Lys Asp Tyr Ser Ser Thr Gly Ile Asn Phe Asp Met Thr Lys
20 25 30

Cys Ile Asn Cys Gln Ser Cys Val Arg Ala Cys Thr Asn Ile Ala Gly
35 40 45

Gln Asn Val Leu Lys Ser Leu Thr Val Asn Gly Lys Ser Val Val Gln
50 55 60

Thr Val Thr Gly Lys Pro Leu Ala Glu Thr Asn Cys Ile Ser Cys Gly
65 70 75 80

Gln Cys Thr Leu Gly Cys Pro Lys Phe Thr Ile Phe Glu Ala Asp Ala
85 90 95

Ile Asn Pro Val Lys Glu Val Leu Thr Lys Lys Asn Gly Arg Ile Ala
100 105 110

Val Cys Gln Ile Ala Pro Ala Ile Arg Ile Asn Met Ala Glu Ala Leu
115 120 125

Gly Val Pro Ala Gly Thr Ile Ser Leu Gly Lys Val Val Thr Ala Leu
130 135 140

Lys Arg Leu Gly Phe Asp Tyr Val Phe Asp Thr Asn Phe Ala Ala Asp
145 150 155 160

Met Thr Ile Val Glu Glu Ala Thr Glu Leu Val Gln Arg Leu Ser Asp
165 170 175

Lys Asn Ala Val Leu Pro Met Phe Thr Ser Cys Cys Pro Ala Trp Val

Asn Tyr Val Glu Lys Ser Asp Pro Ser Leu Ile Pro Tyr Leu Ser Ser
195 200 205

Cys Arg Ser Pro Met Ser Met Leu Ser Ser Val Ile Lys Asn Val Phe
210 215 220

Pro Lys Lys Ile Gly Thr Thr Ala Asp Lys Ile Tyr Asn Val Ala Ile
225 230 235 240

Met Pro Cys Thr Arg Lys Lys Asp Glu Ile Gln Arg Ser Gln Phe Thr
245 250 255

Met Lys Asp Gly Lys Gln Glu Thr Gly Ala Val Leu Thr Ser Arg Glu
260 265 270

Leu Ala Lys Met Ile Lys Glu Ala Lys Ile Asn Phe Lys Glu Leu Pro
275 280 285

Asp Thr Pro Cys Asp Asn Phe Tyr Ser Glu Ala Ser Gly Gly Gly Ala
290 295 300

Ile Phe Cys Ala Thr Gly Gly Val Met Glu Ala Ala Val Arg Ser Ala
305 310 315 320

Tyr Lys Phe Leu Thr Lys Lys Glu Leu Ala Pro Ile Asp Leu Gln Asp
325 330 335

Val Arg Gly Val Ala Ser Gly Val Lys Leu Ala Glu Val Asp Ile Ala
340 345 350

Gly Thr Lys Val Lys Val Ala Val Ala His Gly Ile Lys Asn Ala Met
355 360 365

Thr Leu Ile Lys Lys Ile Lys Ser Gly Glu Glu Gln Phe Lys Asp Val
370 375 380

Lys Phe Val Glu Val Met Ala Cys Pro Gly Gly Cys Val Val Gly Gly
385 390 395 400

Gly Ser Pro Lys Ala Lys Thr Lys Lys Ala Val Gln Ala Arg Leu Asn
405 410 415

Ala Thr Tyr Ser Ile Asp Lys Ser Ser Lys His Arg Thr Ser Gln Asp
420 425 430

Asn Pro Gln Leu Leu Gln Leu Tyr Lys Glu Ser Phe Glu Gly Lys Phe
435 440 445

Gly Gly His Val Ala His His Leu Leu His Thr His Tyr Lys Asn Arg
450 455 460

Lys Val Asn Pro
465

<210> 20
<211> 449
<212> PRT
<213> Trichomonas vaginalis

<400> 21

Met Leu Ala Ser Ser Ser Arg Ala Ala Ala Asn Ile Arg Trp Val Asp
1 5 10 15

Thr Ser His Asn Ala Ile Ala Phe Asp Met His Lys Cys Ile Asn Cys
20 25 30

Gln Ala Cys Val Arg Ala Cys Lys Asn Val Ala Gly Gln Ser Val Leu
35 40 45

Lys Ser Val Lys Ile Asn Glu Gly Lys Lys Lys Gly Val Val Gln Thr
50 55 60

Val Thr Gly Lys Leu Leu Ala Glu Thr Asn Cys Ile Gly Cys Gly Gln
65 70 75 80

Cys Thr Leu Val Cys Pro Thr Gln Ala Ile His Glu Lys Asp Ala Leu
85 90 95

Lys Gln Met Asn Asn Ile Phe Lys Asn Lys Gly Asp Arg Ile Leu Val
100 105 110

Cys Gln Ile Ala Pro Ala Ile Arg Ile Asn Met Arg Arg Pro Trp Cys
115 120 125

Ser Ser Arg Asn Ser Phe His Arg Gln Ser Arg Tyr Ser Pro Gln Arg
130 135 140

Leu Gly Phe Asp Tyr Val Phe Asp Thr Asn Phe Gly Ala Asp Leu Thr
145 150 155 160

Ile Val Glu Glu Ala Thr Glu Leu Leu Gln Arg Leu Asn Asp Pro Lys
165 170 175

Ala Val Leu Pro Met Phe Thr Ser Cys Cys Pro Ala Trp Val Asn Tyr
180 185 190

Val Glu Lys Ser Tyr Pro Gln Trp Met Pro His Leu Ser Thr Cys Arg
195 200 205

Ser Pro Ile Gly Met Leu Ser Ala Val Ile Lys Asn Val Phe Pro Lys
210 215 220

His Ile Gly Val Asp Pro Lys Arg Ile Phe Ser Val Gly Ile Met Pro
225 230 235 240

Cys Thr Ala Lys Lys Asp Glu Ala Ala Arg Glu Gln Leu Met Thr Lys
245 250 255

Ser Gly Leu His Glu Thr Asp Leu Asp Ile Thr Ser Arg Glu Leu Ala
260 265 270

PCT/US2005/001983

Lys Met Ile Lys Ala Ala Lys Ile Asn Phe Lys Glu Leu Pro Asp Thr
275 280 285

Glu Leu Asp Ser Pro Tyr Ala Met Ala Thr Gly Gly Gly Ala Ile Phe
290 295 300

Cys Ala Thr Gly Gly Val Met Glu Ala Ala Val Arg Ser Ala Tyr Lys
305 310 315 320

Phe Ala Thr Gly Lys Glu Leu Ala Pro Ile Glu Phe Val Gln Val Arg
325 330 335

Gly Ala Glu Lys Gly Ile Lys Val Gly Thr Val Asp Ile Asn Gly Arg
340 345 350

Glu Ile Lys Val Ala Val Ala Gln Gly Val Lys Asn Ala Met Ser Leu
355 360 365

Ile Lys Lys Ile Glu Glu Gly Gln Asp Asp Val Lys Gly Val Val Phe
370 375 380

Cys Glu Val Met Ala Cys Pro Gly Gly Cys Val Gly Gly Gly Gly Ser
385 390 395 400

Pro Arg Ala Lys Thr Lys Ala Ala Met Asn Lys Arg Leu Asp Ala Thr
405 410 415

Tyr Arg Ile Asp Arg Ala Ser Lys Tyr Arg Thr Pro Gln Asp Asn Thr
420 425 430

Gln Leu Gln Asp Leu Tyr Asn Ala Thr Trp Val Val Ser Leu Val Met
435 440 445

Asp

<210> 22
<211> 589
<212> PRT
<213> Trichomonas vaginalis

<400> 22

Ala Ser Thr Gly Ile Asn Ser Thr Ala Asn Ile Leu Arg Asn Ile Thr
1 5 10 15

Val Thr Val Asn Gly Lys Pro Leu Glu Ala Lys Lys Gly Glu Thr Val
20 25 30

Leu Glu Leu Cys Asp Arg Asn Asn Ile Arg Ile Pro Arg Leu Cys Phe
35 40 45

His Pro Asn Leu Pro Pro Lys Ala Ser Cys Arg Val Cys Leu Val Glu
50 55 60

Cys Asp Gly Lys Trp Leu Ser Pro Ala Cys Val Thr Thr Val Trp Asp
65 70 75 80

Gly Leu Lys Ile Asp Thr Lys Ser Lys Asn Val Arg Asp Ser Val Glu
85 90 95

Asn Asn Leu Lys Glu Leu Leu Asp Cys His Asp Glu Thr Cys Ser Ala
100 105 110

Cys Ile Ala Asn His Arg Cys Gln Phe Arg Asp Met Asn Val Ala Tyr
115 120 125

Ser Val Lys Ala Glu Thr Lys Glu Ile Cys Ser Glu Glu Gly Ile Asp
130 135 140

Glu Ser Thr Asn Ala Ile Arg Leu Asp Thr Ser Lys Cys Val Leu Cys
145 150 155 160

Gly Arg Cys Ile Arg Ala Cys Glu Glu Val Ala Gly Thr Ser Ala Ile
165 170 175

Ile Phe Gly Asn Arg Ala Lys Lys Met Arg Ile Gln Pro Thr Phe Gly
180 185 190

Val Thr Leu Gln Glu Thr Ser Cys Ile Lys Cys Gly Gln Cys Thr Leu
195 200 205

Tyr Cys Pro Val Gly Ala Ile Thr Glu Lys Ser Gln Val Lys Glu Ala
210 215 220

Leu Asp Ile Leu Ala Asn Lys Gly Lys Lys Ile Thr Val Val Gln Val
225 230 235 240

Ala Pro Ala Val Arg Val Ala Leu Ser Glu Ala Phe Gly Tyr Lys Glu
245 250 255

Gly Thr Val Thr Thr Gly Lys Met Val Ser Ala Leu Lys Ala Leu Gly
260 265 270

Phe Asp Leu Val Tyr Asp Thr Asn Tyr Gly Ala Asp Leu Thr Ile Cys
275 280 285

Glu Glu Ala Gly Glu Leu Val Asn Arg Leu Arg Asp Pro Asn Ala Lys
290 295 300

Phe Pro Met Phe Thr Thr Cys Cys Pro Ala Trp Val Asn Tyr Val Glu
305 310 315 320

Gln Ser Ala Pro Asp Phe Ile Pro Asn Leu Ser Ser Cys Arg Ser Pro
325 330 335

Gln Gly Met Leu Ser Ala Leu Ile Lys Asn Tyr Leu Pro Lys Leu Leu
340 345 350

Asp Val Lys Gln Glu Asp Val Leu Asn Phe Ser Ile Met Pro Cys Thr
355 360 365

Ala Lys Lys Asp Glu Val Glu Arg Pro Glu Leu Arg Thr Lys Ser Gly
370 375 380

Leu Lys Glu Thr Asp Met Val Leu Thr Val Arg Glu Leu Val Glu Met
385 390 395 400

Ile Lys Leu Ser Asn Ile Asp Phe Asn Asn Leu Pro Asp Thr Gln Phe
405 410 415

Asp Asn Ile Phe Gly Phe Gly Ser Gly Ala Gly Gln Ile Phe Ala Ala
420 425 430

Thr Gly Gly Val Met Glu Ala Ala Ser Arg Thr Ala Phe Glu Val Tyr
435 440 445

Thr Gly Lys Lys Leu Thr Asn Val Asn Ile Tyr Pro Val Arg Gly Met
450 455 460

Asp Gly Leu Arg Ile Ala Glu Leu Asp Leu Asp Gly Thr Lys Leu Lys
465 470 475 480

Val Ala Val Cys His Gly Ile Ala Asn Thr Ala Lys Leu Leu Asp Arg
485 490 495

Leu Arg Glu Lys Asp Pro Glu Leu Met Asp Ile Lys Phe Ile Glu Ile
500 505 510

Met Ala Cys Pro Gly Gly Cys Val Cys Gly Gly Gly Thr Pro Gln Pro
515 520 525

Lys Asn Arg Val Ser Leu Asp Asn Arg Leu Ala Ala Ile Tyr Asn Ile
530 535 540

Asp Ala Lys Met Glu Cys Arg Lys Ser His Glu Asn Pro Leu Ile Lys
545 550 555 560

Gly Val Tyr Lys Glu Phe Leu Gly Lys Pro Asn Ser His Leu Ala His
565 570 575

Glu Leu Leu His Thr His Phe Lys His His Pro Lys Trp
580 585

<210> 23
<211> 582
<212> PRT
<213> Trichomonas vaginalis

<400> 23

Met Lys Thr Ile Ile Leu Asn Gly Asn Glu Val His Thr Asp Lys Asp
1 5 10 15

Ile Thr Ile Leu Glu Leu Ala Arg Glu Asn Asn Val Asp Ile Pro Thr
20 25 30

Leu Cys Phe Leu Lys Asp Cys Gly Asn Phe Gly Lys Cys Gly Val Cys
35 40 45

Met Val Glu Val Glu Gly Lys Gly Phe Arg Ala Ala Cys Val Ala Lys
50 55 60

Val Glu Asp Gly Met Val Ile Asn Thr Glu Ser Asp Glu Val Lys Glu
65 70 75 80

Arg Ile Lys Lys Arg Val Ser Met Leu Leu Asp Lys His Glu Phe Lys
85 90 95

Cys Gly Gln Cys Ser Arg Arg Glu Asn Cys Glu Phe Leu Lys Leu Val
100 105 110

Ile Lys Thr Lys Ala Lys Ala Ser Lys Pro Phe Leu Pro Glu Asp Lys
115 120 125

Asp Ala Leu Val Asp Asn Arg Ser Lys Ala Ile Val Ile Asp Arg Ser
130 135 140

Lys Cys Val Leu Cys Gly Arg Cys Val Ala Ala Cys Lys Gln His Thr
145 150 155 160

Ser Thr Cys Ser Ile Gln Phe Ile Lys Lys Asp Gly Gln Arg Ala Val
165 170 175

Gly Thr Val Asp Asp Val Cys Leu Asp Asp Ser Thr Cys Leu Leu Cys
180 185 190

Gly Gln Cys Val Ile Ala Cys Pro Val Ala Ala Leu Lys Glu Lys Ser
195 200 205

His Ile Glu Lys Val Gln Glu Ala Leu Asn Asp Pro Lys Lys His Val
210 215 220

Ile Val Ala Met Ala Pro Ser Val Arg Thr Ala Met Gly Glu Leu Phe
225 230 235 240

Lys Met Gly Tyr Gly Lys Asp Val Thr Gly Lys Leu Tyr Thr Ala Leu
245 250 255

Arg Met Leu Gly Phe Asp Lys Val Phe Asp Ile Asn Phe Gly Ala Asp
260 265 270

Met Thr Ile Met Glu Glu Ala Thr Glu Leu Leu Gly Arg Val Lys Asn
275 280 285

Asn Gly Pro Phe Pro Met Phe Thr Ser Cys Cys Pro Ala Trp Val Arg
290 295 300

Leu Ala Gln Asn Tyr His Pro Glu Leu Leu Asp Asn Leu Ser Ser Ala
305 310 315 320

Lys Ser Pro Gln Gln Ile Phe Gly Thr Ala Ser Lys Thr Tyr Tyr Pro
325 330 335

Ser Ile Ser Gly Ile Ala Pro Glu Asp Val Tyr Thr Val Thr Ile Met
340 345 350

Pro Cys Asn Asp Lys Lys Tyr Glu Ala Asp Ile Pro Phe Met Glu Thr
355 360 365

```
Asn Ser Leu Arg Asp Ile Asp Ala Ser Leu Thr Thr Arg Glu Leu Ala
370                 375                 380

Lys Met Ile Lys Asp Ala Lys Ile Lys Phe Ala Asp Leu Glu Asp Gly
385                 390                 395                 400

Glu Val Asp Pro Ala Met Gly Thr Tyr Ser Gly Ala Gly Ala Ile Phe
                405                 410                 415

Gly Ala Thr Gly Gly Val Met Glu Ala Ala Ile Arg Ser Ala Lys Asp
            420                 425                 430

Phe Ala Glu Asn Lys Glu Leu Glu Asn Val Asp Tyr Thr Glu Val Arg
        435                 440                 445

Gly Phe Lys Gly Ile Lys Glu Ala Glu Val Glu Ile Ala Gly Asn Lys
    450                 455                 460

Leu Asn Val Ala Val Ile Asn Gly Ala Ser Asn Phe Phe Glu Phe Met
465                 470                 475                 480

Lys Ser Gly Lys Met Asn Glu Lys Gln Tyr His Phe Ile Glu Val Met
                485                 490                 495

Ala Cys Pro Gly Gly Cys Ile Asn Gly Gly Gly Gln Pro His Val Asn
            500                 505                 510

Ala Leu Asp Arg Glu Asn Val Asp Tyr Arg Lys Leu Arg Ala Ser Val
        515                 520                 525

Leu Tyr Asn Gln Asp Lys Asn Val Leu Ser Lys Arg Lys Ser His Asp
    530                 535                 540

Asn Pro Ala Ile Ile Lys Met Tyr Asp Ser Tyr Phe Gly Lys Pro Gly
545                 550                 555                 560

Glu Gly Leu Ala His Lys Leu Leu His Val Lys Tyr Thr Lys Asp Lys
                565                 570                 575

Asn Val Ser Lys His Glu
            580
```

<210> 24
<211> 497
<212> PRT
<213> Chlamydomonas reinhardtii

<400> 24

```
Met Ser Ala Leu Val Leu Lys Pro Cys Ala Ala Val Ser Ile Arg Gly
1               5                   10                  15

Ser Ser Cys Arg Ala Arg Gln Val Ala Pro Arg Ala Pro Leu Ala Ala
            20                  25                  30

Ser Thr Val Arg Val Ala Leu Ala Thr Leu Glu Ala Pro Ala Arg Arg
        35                  40                  45
```

Leu Gly Asn Val Ala Cys Ala Ala Ala Ala Pro Ala Ala Glu Ala Pro
50 55 60

Leu Ser His Val Gln Gln Ala Leu Ala Glu Leu Ala Lys Pro Lys Asp
65 70 75 80

Asp Pro Thr Arg Lys His Val Cys Val Gln Val Ala Pro Ala Val Arg
85 90 95

Val Ala Ile Ala Glu Thr Leu Gly Leu Ala Pro Gly Ala Thr Thr Pro
100 105 110

Lys Gln Leu Ala Glu Gly Leu Arg Arg Leu Gly Phe Asp Glu Val Phe
115 120 125

Asp Thr Leu Phe Gly Ala Asp Leu Thr Ile Met Glu Glu Gly Ser Glu
130 135 140

Leu Leu His Arg Leu Thr Glu His Leu Glu Ala His Pro His Ser Asp
145 150 155 160

Glu Pro Leu Pro Met Phe Thr Ser Cys Cys Pro Gly Trp Ile Ala Met
165 170 175

Leu Glu Lys Ser Tyr Pro Asp Leu Ile Pro Tyr Val Ser Ser Cys Lys
180 185 190

Ser Pro Gln Met Met Leu Ala Ala Met Val Lys Ser Tyr Leu Ala Glu
195 200 205

Lys Lys Gly Ile Ala Pro Lys Asp Met Val Met Val Ser Ile Met Pro
210 215 220

Cys Thr Arg Lys Gln Ser Glu Ala Asp Arg Asp Trp Phe Cys Val Asp
225 230 235 240

Ala Asp Pro Thr Leu Arg Gln Leu Asp His Val Ile Thr Thr Val Glu
245 250 255

Leu Gly Asn Ile Phe Lys Glu Arg Gly Ile Asn Leu Ala Glu Leu Pro
260 265 270

Glu Gly Glu Trp Asp Asn Pro Met Gly Val Gly Ser Gly Ala Gly Val
275 280 285

Leu Phe Gly Thr Thr Gly Gly Val Met Glu Ala Ala Leu Arg Thr Ala
290 295 300

Tyr Glu Leu Phe Thr Gly Thr Pro Leu Pro Arg Leu Ser Leu Ser Glu
305 310 315 320

Val Arg Gly Met Asp Gly Ile Lys Glu Thr Asn Ile Thr Met Val Pro
325 330 335

Ala Pro Gly Ser Lys Phe Glu Glu Leu Leu Lys His Arg Ala Ala Ala

340 345 350

Arg Ala Glu Ala Ala Ala His Gly Thr Pro Gly Pro Leu Ala Trp Asp
355 360 365

Gly Gly Ala Gly Phe Thr Ser Glu Asp Gly Arg Gly Gly Ile Thr Leu
370 375 380

Arg Val Ala Val Ala Asn Gly Leu Gly Asn Ala Lys Lys Leu Ile Thr
385 390 395 400

Lys Met Gln Ala Gly Glu Ala Lys Tyr Asp Phe Val Glu Ile Met Ala
405 410 415

Cys Pro Ala Gly Cys Val Gly Gly Gly Gly Gln Pro Arg Ser Thr Asp
420 425 430

Lys Ala Ile Thr Gln Lys Arg Gln Ala Ala Leu Tyr Asn Leu Asp Glu
435 440 445

Lys Ser Thr Leu Arg Arg Ser His Glu Asn Pro Ser Ile Arg Glu Leu
450 455 460

Tyr Asp Thr Tyr Leu Gly Glu Pro Leu Gly His Lys Ala His Glu Leu
465 470 475 480

Leu His Thr His Tyr Val Ala Gly Gly Val Glu Glu Lys Asp Glu Lys
485 490 495

Lys

<210> 25
<211> 415
<212> PRT
<213> Chlorella fusca

<400> 25

Ala Gly Pro Thr Ser Glu Cys Asp Cys Pro Pro Thr Pro Gln Ala Lys
1 5 10 15

Leu Pro His Trp Gln Gln Ala Leu Asp Glu Leu Ala Lys Pro Lys Glu
20 25 30

Ser Arg Arg Leu Met Ile Ala Gln Ile Ala Ser Ala Val Arg Val Ala
35 40 45

Ile Ala Glu Thr Ile Gly Leu Ala Pro Gly Asp Val Thr Ile Gly Gln
50 55 60

Leu Val Thr Gly Leu Arg Met Leu Gly Phe Asp Tyr Val Phe Asp Thr
65 70 75 80

Leu Phe Gly Ala Asp Leu Thr Ile Met Glu Glu Gly Thr Glu Leu Leu
85 90 95

His Arg Leu Gln Asp His Leu Glu Gln His Pro Asn Lys Glu Glu Pro

100 105 110

Leu Pro Met Phe Thr Ser Cys Cys Pro Gly Trp Val Ala Met Val Glu
115 120 125

Lys Ser Asn Pro Glu Leu Ile Pro Tyr Leu Ser Ser Cys Lys Ser Pro
130 135 140

Gln Met Met Leu Gly Ala Val Ile Lys Asn Tyr Tyr Ala Gln Gln Val
145 150 155 160

Gly Val Gln Pro Ser Asp Ile Cys Asn Val Ser Val Met Pro Cys Val
165 170 175

Arg Lys Gln Gly Glu Ala Asp Arg Glu Trp Phe Asn Thr Thr Gly Ala
180 185 190

Gly Leu Ala Arg Asp Val Asp His Val Val Thr Thr Ala Glu Val Gly
195 200 205

Lys Ile Phe Leu Glu Arg Gly Ile Lys Leu Asn Glu Leu Pro Glu Ser
210 215 220

Asn Phe Asp Asn Pro Ile Gly Glu Gly Thr Gly Gly Ala Leu Leu Phe
225 230 235 240

Gly Thr Thr Gly Gly Val Met Glu Ala Ala Leu Arg Thr Val Tyr Glu
245 250 255

Val Val Thr Gln Lys Pro Met Gly Arg Val Asp Phe Glu Glu Val Arg
260 265 270

Gly Leu Glu Gly Ile Lys Glu Ala Glu Ile Thr Leu Lys Pro Gly Asp
275 280 285

Asp Ser Pro Phe Lys Ala Phe Ala Gly Ala Asp Gly Gln Gly Ile Thr
290 295 300

Leu Lys Ile Ala Val Ala Asn Gly Leu Gly Asn Ala Lys Lys Leu Ile
305 310 315 320

Lys Ser Leu Ser Glu Gly Lys Ala Lys Tyr Asp Phe Ile Glu Val Met
325 330 335

Ala Cys Pro Gly Gly Cys Ile Gly Gly Gly Gly Gln Pro Arg Ser Thr
340 345 350

Asp Lys Gln Ile Leu Gln Lys Arg Gln Gln Ala Met Tyr Asn Leu Asp
355 360 365

Glu Arg Ser Thr Ile Arg Arg Ser His Asp Asn Pro Phe Ile Gln Ala
370 375 380

Leu Tyr Asp Lys Phe Leu Gly Ala Pro Asn Ser His Lys Ala His Asp
385 390 395 400

Leu Leu His Thr His Tyr Val Ala Gly Gly Ile Pro Glu Glu Lys
405 410 415

<210> 26
<211> 505
<212> PRT
<213> Chlamydomonas reinhardtii

<400> 26

Met Ala Leu Gly Leu Leu Ala Glu Leu Arg Ala Gly Gln Ala Val Ala
1 5 10 15

Cys Ala Arg Arg Thr Asn Ala Pro Ala His Pro Ala Ala Val Val Pro
20 25 30

Cys Leu Pro Ser Arg Ala Gly Lys Phe Phe Asn Leu Ser Gln Lys Val
35 40 45

Pro Ser Ser Gln Ser Ala Arg Gly Ser Thr Ile Arg Val Ala Ala Thr
50 55 60

Ala Thr Asp Ala Val Pro His Trp Lys Leu Ala Leu Glu Glu Leu Asp
65 70 75 80

Lys Pro Lys Asp Gly Gly Arg Lys Val Leu Ile Ala Gln Val Ala Pro
85 90 95

Ala Val Arg Val Ala Ile Ala Glu Ser Phe Gly Leu Ala Pro Gly Ala
100 105 110

Val Ser Pro Gly Lys Leu Ala Thr Gly Leu Arg Ala Leu Gly Phe Asp
115 120 125

Gln Val Phe Asp Thr Leu Phe Ala Ala Asp Leu Thr Ile Met Glu Glu
130 135 140

Gly Thr Glu Leu Leu His Arg Leu Lys Glu His Leu Glu Ala His Pro
145 150 155 160

His Ser Asp Glu Pro Leu Pro Met Phe Thr Ser Cys Cys Pro Gly Trp
165 170 175

Val Ala Met Met Glu Lys Ser Tyr Pro Glu Leu Ile Pro Phe Val Ser
180 185 190

Ser Cys Lys Ser Pro Gln Met Met Met Gly Ala Met Val Lys Thr Tyr
195 200 205

Leu Ser Glu Lys Gln Gly Ile Pro Ala Lys Asp Ile Val Met Val Ser
210 215 220

Val Met Pro Cys Val Arg Lys Gln Gly Glu Ala Asp Arg Glu Trp Phe
225 230 235 240

Cys Val Ser Glu Pro Gly Val Arg Asp Val Asp His Val Ile Thr Thr
245 250 255

Ala Glu Leu Gly Asn Ile Phe Lys Glu Arg Gly Ile Asn Leu Pro Glu
260 265 270

Leu Pro Asp Ser Asp Trp Asp Gln Pro Leu Gly Leu Gly Ser Gly Ala
275 280 285

Gly Val Leu Phe Gly Thr Thr Gly Gly Val Met Glu Ala Ala Leu Arg
290 295 300

Thr Ala Tyr Glu Ile Val Thr Lys Glu Pro Leu Pro Arg Leu Asn Leu
305 310 315 320

Ser Glu Val Arg Gly Leu Asp Gly Ile Lys Glu Ala Ser Val Thr Leu
325 330 335

Val Pro Ala Pro Gly Ser Lys Phe Ala Glu Leu Val Ala Glu Arg Leu
340 345 350

Ala His Lys Val Glu Glu Ala Ala Ala Ala Glu Ala Ala Ala Ala Val
355 360 365

Glu Gly Ala Val Lys Pro Pro Ile Ala Tyr Asp Gly Gly Gln Gly Phe
370 375 380

Ser Thr Asp Asp Gly Lys Gly Gly Leu Lys Leu Arg Val Ala Val Ala
385 390 395 400

Asn Gly Leu Gly Asn Ala Lys Lys Leu Ile Gly Lys Met Val Ser Gly
405 410 415

Glu Ala Lys Tyr Asp Phe Val Glu Ile Met Ala Cys Pro Ala Gly Cys
420 425 430

Val Gly Gly Gly Gly Gln Pro Arg Ser Thr Asp Lys Gln Ile Thr Gln
435 440 445

Lys Arg Gln Ala Ala Leu Tyr Asp Leu Asp Glu Arg Asn Thr Leu Arg
450 455 460

Arg Ser His Glu Asn Glu Ala Val Asn Gln Leu Tyr Lys Glu Phe Leu
465 470 475 480

Gly Glu Pro Leu Ser His Arg Ala His Glu Leu Leu His Thr His Tyr
485 490 495

Val Pro Gly Gly Ala Glu Ala Asp Ala
500 505

<210> 27
<211> 403
<212> PRT
<213> Scenedesmus obliquus

<400> 27

Pro His Trp Gln Gln Thr Leu Asp Glu Leu Ala Lys Pro Lys Glu Arg
1 5 10 15

Lys Val Met Ile Ala Gln Ile Ala Pro Ala Val Arg Gly Ile Ala Glu
20 25 30

Thr Met Gly Leu Asn Pro Gly Asp Val Thr Val Gly Gln Met Val Thr
35 40 45

Gly Leu Arg Met Leu Gly Phe Asp Tyr Val Phe Asp Thr Leu Phe Gly
50 55 60

Ala Asp Leu Thr Ile Met Glu Glu Gly Thr Glu Leu Leu His Arg Leu
65 70 75 80

Gln Asp His Leu Glu Gln His Pro Asn Lys Glu Glu Pro Leu Pro Met
85 90 95

Phe Thr Ser Cys Cys Pro Gly Trp Val Ala Met Val Glu Lys Ser Asn
100 105 110

Pro Glu Leu Ile Pro Tyr Leu Ser Ser Cys Lys Ser Pro Gln Met Met
115 120 125

Leu Gly Ala Val Ile Lys Asn Tyr Phe Ala Ala Glu Ala Gly Ala Lys
130 135 140

Pro Glu Asp Ile Cys Asn Val Ser Val Met Pro Cys Val Arg Lys Ser
145 150 155 160

Gly Glu Ala Glu Pro Arg Ser Gly Ser Thr His His Arg Ala Gly Arg
165 170 175

Arg Asp Val Asp His Val Met Thr Thr Ala Glu Leu Gly Lys Ile Phe
180 185 190

Val Glu Arg Gly Ile Lys Leu Asn Glu Leu Gln Glu Ser Pro Phe Asp
195 200 205

Asn Pro Val Gly Glu Gly Ser Gly Gly Gly Leu Leu Phe Gly Thr Thr
210 215 220

Gly Gly Val Met Glu Ala Ala Leu Arg Thr Val Tyr Glu Val Val Thr
225 230 235 240

Ala Glu Ala Leu Gly Pro Gln Arg Ser Ser Leu Thr Thr Ser Thr Ala
245 250 255

Trp Thr Pro Ala Gln Arg Ala Ser Pro Arg Pro Ser Pro Gln Ala Pro
260 265 270

Thr Ala Pro Ser Arg Pro Leu Gln Ala Gln Thr Glu Ser Gly Ile Thr
275 280 285

Leu Asn Ile Ala Val Ala Asn Gly Leu Gly Asn Ala Lys Lys Leu Ile
290 295 300

Lys Gln Leu Ala Ala Gly Glu Ser Lys Tyr Asp Phe Thr Glu Val Met
305 310 315 320

Ala Cys Pro Gly Gly Cys Ile Gly Gly Gly Gly Gln Pro Gln Arg Asn
325 330 335

Lys Gln Ile Leu Gln Lys Arg Gln Ala Ala Met Tyr Asp Leu Asp Glu
340 345 350

Arg Ala Val Ile Arg Arg Thr Glu Asn Pro Leu Ile Gly Ala Leu Tyr
355 360 365

Glu Lys Phe Leu Gly Glu Pro Asn Gly His Lys Ala His Glu Leu Leu
370 375 380

His Thr His Tyr Val Ala Gly Gly Val Pro Asp Arg Arg Ser Glu Gly
385 390 395 400

Glu Ala Trp

<210> 28
<211> 581
<212> PRT
<213> Thermoanaerobacter tengcongensis strain MB4T

<400> 28

Met Asp Lys Val Arg Val Thr Ile Asp Gly Ile Thr Val Glu Val Pro
1 5 10 15

Ser Tyr Tyr Thr Val Leu Glu Ala Ala Lys Glu Ala Gly Ile Asp Ile
20 25 30

Pro Thr Leu Cys Tyr Leu Lys Glu Ile Asn Gln Ile Gly Ala Cys Arg
35 40 45

Ile Cys Leu Val Glu Ile Glu Gly Val Arg Asn Leu Gln Thr Ser Cys
50 55 60

Thr Tyr Pro Val Phe Asp Gly Met Lys Val Tyr Thr Asn Thr Pro Lys
65 70 75 80

Ile Arg Glu Ala Arg Arg Leu Asn Leu Glu Leu Ile Leu Ser Asn His
85 90 95

Asp Arg Asn Cys Leu Thr Cys Val Arg Ser Thr Asn Cys Glu Leu Gln
100 105 110

Ala Leu Ala Lys Arg Leu Gly Val Glu Glu Ile Arg Phe Glu Gly Glu
115 120 125

Asn Ile Lys Tyr Pro Ile Asp Asp Ala Ser Pro Ala Val Val Arg Asp
130 135 140

Pro Asn Lys Cys Val Leu Cys Arg Arg Cys Val Ala Val Cys Ser Glu
145 150 155 160

Val Gln Asn Val Phe Ala Ile Gly Met Val Asn Arg Gly Phe Lys Thr
165 170 175

Met Val Ala Pro Ser Phe Gly Arg Ser Leu Lys Asp Ser Pro Cys Ile
180 185 190

Ser Cys Gly Gln Cys Ile Met Val Cys Pro Val Gly Ala Ile Tyr Glu
195 200 205

Lys Asp His Thr Lys Arg Val Tyr Glu Ala Leu Ala Asp Asp Lys Lys
210 215 220

Tyr Val Val Ala Gln Thr Ala Pro Ala Val Arg Val Ala Leu Gly Glu
225 230 235 240

Glu Phe Gly Met Pro Val Gly Thr Ile Val Thr Gly Lys Met Ala Ala
245 250 255

Ala Leu Arg Arg Met Gly Phe Asp Ala Val Phe Asp Thr Asn Phe Ala
260 265 270

Ala Asp Leu Thr Ile Met Glu Glu Gly Ser Glu Leu Leu Glu Arg Ile
275 280 285

Lys His Gly Gly Lys Leu Pro Met Ile Thr Ser Cys Ser Pro Gly Trp
290 295 300

Ile Ala Phe Cys Glu Lys Tyr Tyr Pro Glu Phe Ile Asp Asn Leu Ser
305 310 315 320

Thr Cys Lys Ser Pro His Met Met Met Gly Ala Leu Val Lys Ser Tyr
325 330 335

Tyr Ala Glu Lys Lys Gly Leu Asp Pro Lys Asp Ile Phe Val Val Ser
340 345 350

Ile Met Pro Cys Thr Ala Lys Lys Leu Glu Ile Glu Arg Glu Glu Met
355 360 365

Ile Arg Asn Gly Met Lys Asp Val Asp Ala Val Leu Thr Thr Arg Glu
370 375 380

Leu Ala Arg Met Ile Lys Glu Met Gly Ile Asp Phe Val Asn Leu Lys
385 390 395 400

Asp Glu Glu Phe Asp Glu Pro Leu Gly Met Ser Thr Gly Ala Gly Ala
405 410 415

Ile Phe Gly Ala Thr Gly Gly Val Met Glu Ala Ala Leu Arg Thr Val
420 425 430

Ala Glu Ile Val Glu Gly Arg Asp Ile Gly Lys Ile Asp Phe Glu Glu
435 440 445

Val Arg Gly Leu Glu Gly Val Arg Glu Ala Thr Ile Thr Ile Asp Gly
450 455 460

Met Asp Ile Lys Ile Ala Ile Ala Asn Gly Thr Gly Asn Ala Lys Lys
465 470 475 480

Leu Leu Asp Lys Val Lys Ala Gly Glu Val Glu Tyr His Phe Ile Glu
485 490 495

Val Met Gly Cys Pro Gly Gly Cys Ile Met Gly Gly Gly Gln Pro Ile
500 505 510

His Asn Pro Asn Glu Met Glu Glu Val Lys Lys Leu Arg Ala Lys Ala
515 520 525

Ile Tyr Glu Ile Asp Lys Asn Leu Pro Ile Arg Lys Ser His Glu Asn
530 535 540

Pro Ala Ile Lys Arg Leu Tyr Glu Glu Phe Leu Gly Tyr Pro Leu Ser
545 550 555 560

Glu Lys Ser His Glu Leu Leu His Thr His Tyr Ser Arg Lys Glu Leu
565 570 575

Tyr Pro Leu Val Lys
580

<210> 29
<211> 636
<212> PRT
<213> Neocallimastix frontalis

<400> 29

Met Ser Met Leu Ser Ser Val Leu Asn Lys Ala Val Val Asn Pro Lys
1 5 10 15

Leu Thr Arg Ser Leu Ala Thr Ala Ala Ala Glu Lys Met Val Asn Ile
20 25 30

Ser Ile Asn Gly Arg Lys Phe Gln Val Lys Pro Lys Thr Thr Val Leu
35 40 45

Glu Ala Ala Lys Ala Asn Gly Tyr Tyr Ile Pro Thr Leu Cys Tyr His
50 55 60

Gln Glu Leu Pro Val Ala Gly Asn Cys Arg Leu Cys Leu Val Tyr Ala
65 70 75 80

Lys Gly Ser Trp Lys Pro Leu Thr Ala Cys Thr Thr Glu Val Trp Glu
85 90 95

Gly Met Glu Ile Glu Thr Asp Ser Pro Ala Val Ile Glu Thr Val Arg
100 105 110

Ser Ser Leu Ser Met Met Arg Glu Glu His Pro Asn Asp Cys Met Thr
115 120 125

Cys Gly Ser Asn Gly Asp Cys Glu Phe Gln Asp Leu Ile Tyr Arg Tyr
130 135 140

Gln Ile Asp Ala Lys His Pro Val Arg Ser Leu Leu Lys His Lys Ser
145 150 155 160

Lys Lys Thr Asn His Ser Ile Thr Glu Pro Cys Tyr Ser Pro Phe Asp
165 170 175

Asn Thr Thr Phe Ser Val Ala Arg Asp Met Asn Lys Cys Val Lys Cys
180 185 190

Gly Arg Cys Ile Arg Ala Cys His His Phe Gln Asn Ile Asn Ile Leu
195 200 205

Gly Phe Ile Asn Arg Ala Gly Tyr Glu Arg Val Gly Thr Pro Met Asp
210 215 220

Arg Pro Met Asn Phe Thr Lys Cys Val Glu Cys Gly Gln Cys Ser Gln
225 230 235 240

Val Cys Pro Val Gly Ala Ile Thr Ala Arg Thr Glu Val Val Asp Val
245 250 255

Leu Arg His Leu Asp Thr Lys Arg Lys Val Val Val Cys Ser Thr Ala
260 265 270

Pro Ala Ile Arg Val Ala Pro Ala Glu Glu Phe Ser Thr Glu Ala Asp
275 280 285

Phe Asp Phe Thr Gly Lys Met Val Ala Gly Leu Arg Lys Leu Gly Phe
290 295 300

Asp Tyr Ile Phe Asp Thr Asn Phe Ser Ala Asp Leu Thr Ile Met Glu
305 310 315 320

Glu Gly Thr Glu Leu Ile Asp Arg Leu Asn Asn Gly Gly Lys Phe Pro
325 330 335

Met Phe Thr Ser Cys Cys Pro Gly Trp Ile Asn Met Val Glu Lys Ser
340 345 350

Tyr Pro Glu Leu Ser Asp Asn Leu Ser Ser Cys Lys Ser Pro Gln Gln
355 360 365

Met Ile Gly Ala Val Ile Lys Ser Tyr Phe Ala Lys Lys Leu Gly Leu
370 375 380

Ser Thr Glu Asp Ile Ile His Val Ser Ile Met Pro Cys Thr Ala Lys
385 390 395 400

Lys Gly Glu Ala Arg Arg Pro Glu Phe Val Gln Lys Gly Lys Asp Gly
405 410 415

Lys Asp Tyr Pro Asp Ile Asp Tyr Val Ile Thr Thr Arg Glu Leu Leu
420 425 430

Thr Leu Leu Lys Leu Lys Lys Ile Asn Pro Ala Glu Leu Pro Asp Asp
435 440 445

Lys Phe Asp Ser Pro Leu Gly Ile Gly Ser Ser Ala Gly Asn Leu Phe

450 455 460

Gly Val Thr Gly Gly Val Met Glu Ala Ala Ile Arg Thr Ala Gln Val
465 470 475 480

Ile Thr Gly Val Glu Asn Pro Ile Pro Leu Gly Glu Leu Lys Ala Ile
485 490 495

Arg Gly Leu Asp Gly Ile Lys Ala Ala Asn Val Pro Leu Lys Thr Lys
500 505 510

Asp Gly Lys Glu Val Ser Val Arg Ala Ala Val Val Ser Gly Gly Ala
515 520 525

Asn Ile Gln Lys Phe Leu Glu Lys Ile Lys Asn Lys Glu Leu Glu Phe
530 535 540

Asp Phe Ile Glu Met Met Met Cys Pro Gly Gly Cys Ile Asn Gly Gly
545 550 555 560

Gly Gln Pro Lys Ser Ala Asp Pro Glu Ile Val Ala Lys Lys Met Gln
565 570 575

Arg Met Tyr Thr Met Asp Asp Gln Ala Lys Leu Arg Leu Cys His Glu
580 585 590

Asn Pro Glu Ile Ile Asp Val Tyr Lys Asn Phe Leu Gly Glu Pro Asn
595 600 605

Ser His Leu Ala His Glu Leu Leu His Thr His Tyr Asn Asp Arg Ser
610 615 620

Lys Thr Ile His Asp Met Gly His His Glu Lys Lys
625 630 635

<210> 30
<211> 555
<212> PRT
<213> Piromyces sp. E2

<400> 30

Cys Leu Val Asp Val Lys Gly Ser Trp Lys Pro Leu Thr Ala Cys Thr
1 5 10 15

Thr Glu Val Trp Glu Gly Met Glu Ile Glu Thr Asp Thr Pro Ala Val
20 25 30

Arg Glu Thr Val Arg Ser Ser Leu Ala Met Met Arg Glu Glu His Pro
35 40 45

Asn Asp Cys Met Thr Cys Glu Ser Asn Gly Asn Cys Glu Phe Gln Asp
50 55 60

Leu Ile Tyr Arg Tyr Gln Ile Asp Ala Gln His Pro Val Arg Thr Leu
65 70 75 80

Leu Arg Asn Lys Phe Lys Lys Thr Asn His Ser Ile Thr Glu Pro Cys

"85" 90 95

Tyr Ser Pro Phe Asp Asp Ser Thr Phe Ser Ile Ser Arg Asp Met Asn
100 105 110

Lys Cys Val Lys Cys Gly Arg Cys Val Arg Ala Cys His His Phe Gln
115 120 125

Asn Ile Asn Ile Leu Gly Phe Ile Asn Arg Ala Gly Tyr Glu Arg Val
130 135 140

Gly Thr Pro Met Asp Arg Pro Met Asn Phe Thr Lys Cys Val Glu Cys
145 150 155 160

Gly Gln Cys Ser Gln Val Cys Pro Val Gly Ala Ile Thr Glu Arg Asn
165 170 175

Glu Cys Ile Glu Val Leu Arg His Leu Asp Thr Lys Arg Lys Ile Val
180 185 190

Val Val Ser Thr Ala Pro Ala Ile Arg Val Ala Leu Ala Glu Glu Phe
195 200 205

Asn Ala Glu Pro Asp Phe Asp Phe Thr Gly Lys Met Val Ala Gly Leu
210 215 220

Lys Lys Leu Gly Phe Asp Tyr Ile Phe Asp Thr Asn Phe Ser Ala Asp
225 230 235 240

Leu Thr Ile Met Glu Glu Gly Thr Glu Leu Ile Thr Arg Leu Asn Glu
245 250 255

Gly Gly Lys Phe Pro Met Phe Thr Ser Cys Cys Pro Gly Trp Ile Asn
260 265 270

Met Val Glu Lys Ser Tyr Pro Glu Ile Arg Asp Asn Leu Ser Ser Cys
275 280 285

Lys Ser Pro Gln Gln Met Ile Gly Ala Val Ile Lys Thr Tyr Phe Ala
290 295 300

Lys Lys Ile Asn Ala Lys Pro Glu Asp Ile Ile His Val Ser Val Met
305 310 315 320

Pro Cys Thr Ala Lys Lys Gly Glu Ala Lys Arg Pro Glu Phe Lys Arg
325 330 335

Asp Gly Val Pro Asp Ile Asp His Val Ile Thr Thr Arg Glu Leu Ile
340 345 350

Thr Leu Leu Lys Leu Lys Arg Ile Asn Pro Ser Glu Leu Lys Asn Glu
355 360 365

Lys Phe Asp Ser Pro Leu Gly Ile Gly Ser Ser Ala Gly Asn Leu Phe
370 375 380

Gly Val Thr Gly Gly Val Met Glu Ala Ala Val Arg Thr Ala Gln Ile
385 390 395 400

Ile Thr Gly Val Glu Asn Pro Ile Pro Leu Gly Glu Leu Lys Ala Ile
405 410 415

Arg Gly Leu Asp Gly Ile Lys Ala Ala Ser Val Pro Leu Lys Thr Lys
420 425 430

Asp Gly Lys Asp Val Asn Val Arg Ala Ala Val Val Ser Gly Gly Ala
435 440 445

Asn Ile Gln Lys Phe Leu Glu Lys Leu Lys Lys Lys Glu Leu Glu Phe
450 455 460

Asp Phe Val Glu Met Met Met Cys Pro Gly Gly Cys Ile Asn Gly Gly
465 470 475 480

Gly Gln Pro Lys Ser Ala Asp Pro Lys Val Val Ala Lys Lys Met Glu
485 490 495

Arg Met Tyr Thr Met Asp Asp Gln Ala Ser Leu Arg Leu Ser His Glu
500 505 510

Asn Pro Glu Ile Thr Gln Ile Tyr Lys Glu Phe Leu Lys Glu Pro Asn
515 520 525

Gly His Leu Ser His Glu Leu Leu His Thr His Tyr Asn Asp Arg Ser
530 535 540

Lys Ala Ile Gln Asp Met Ser Leu His Gln Lys
545 550 555

<210> 31
<211> 389
<212> PRT
<213> Neocallimastix frontalis

<400> 31

Thr Glu Arg Asn Glu Val Ile Glu Val Leu Arg Gln Leu Asp Ser Lys
1 5 10 15

Arg Lys Ile Leu Val Cys Ser Thr Ala Pro Ala Ile Arg Val Ala Leu
20 25 30

Ala Glu Glu Phe Asn Ala Asp Pro Asp Phe Asn Phe Thr Gly Lys Met
35 40 45

Val Ala Gly Leu Arg Lys Leu Gly Phe Asp Tyr Ile Phe Asp Thr Asn
50 55 60

Phe Ser Ala Asp Leu Thr Ile Met Glu Glu Gly Thr Glu Leu Ile Asn
65 70 75 80

Arg Leu Asn Asn Gly Gly Lys Phe Pro Met Phe Thr Ser Cys Cys Pro
85 90 95

Gly Trp Ile Asn Met Val Glu Lys Ser Tyr Pro Glu Leu Arg Glu Asn
100 105 110

Leu Ser Thr Cys Lys Ser Pro Gln Gln Met Ile Gly Ala Leu Ile Lys
115 120 125

Ser Tyr Phe Ala Lys Lys Leu Gly Val Ser Thr Glu Asp Ile Ile His
130 135 140

Val Ser Val Met Pro Cys Thr Ala Lys Lys Gly Glu Ala Lys Arg Pro
145 150 155 160

Glu Phe Val Gln Lys Gly Lys Asp Gly Lys Asn Tyr Pro Asp Ile Asp
165 170 175

Tyr Val Leu Thr Thr Arg Glu Leu Leu Thr Leu Met Lys Leu Lys Lys
180 185 190

Val Asn Pro Ala Glu Leu Ala Asp Asp Lys Leu Asp Ser Pro Leu Gly
195 200 205

Ile Ser Ser Ser Ala Gly Asn Leu Phe Gly Val Thr Gly Gly Val Met
210 215 220

Glu Ala Ala Val Arg Thr Ala Gln Ile Ile Thr Gly Val Glu Asn Pro
225 230 235 240

Ile Pro Leu Gly Glu Leu Lys Ala Val Arg Gly Leu Glu Gly Ile Lys
245 250 255

Ala Ala Thr Val Pro Leu Lys Thr Lys Glu Gly Lys Asp Ile Asn Val
260 265 270

Arg Ala Ala Val Val Ser Gly Gly Ala Asn Ile Gln Lys Phe Leu Glu
275 280 285

Lys Ile Lys Asn Lys Glu Val Glu Phe Asp Phe Val Glu Met Met Met
290 295 300

Cys Pro Gly Gly Cys Ile Asn Gly Gly Gly Gln Pro Lys Ser Ala Asp
305 310 315 320

Pro Lys Ile Val Thr Lys Lys Met Gln Arg Met Tyr Thr Met Asp Glu
325 330 335

Gln Ala Thr Leu Arg Leu Ser His Glu Asn Glu Glu Val Lys Gln Ile
340 345 350

Tyr Lys Glu Phe Leu Ile Glu Pro Asn Gly His Leu Ser His Glu Leu
355 360 365

Leu His Thr His Tyr Asn Asp Arg Ser Lys Ala Ile Gln Asp Met Ser
370 375 380

Leu His Glu Lys Lys
385

<210> 32
<211> 458
<212> PRT
<213> Desulfovibrio desulfuricans

<400> 32

Met Asn Gly Gln Gln Asn Val Ile Arg Ile Asp Ser Asp Ile Cys Thr
1 5 10 15

Gly Cys Gly Arg Cys Lys Asp Val Cys Pro Val Gly Ala Val Glu Gly
20 25 30

Val Gln Gly Thr Pro His Ser Ile Arg Glu Asp Val Cys Val Leu Cys
35 40 45

Gly Gln Cys Val Gln Gln Cys Ser Ala Phe Ala Ser Phe Tyr Glu Gln
50 55 60

His Pro Ala Cys Ile Ala Glu Lys Lys Arg Glu Arg Gly Leu Phe Val
65 70 75 80

Ser Glu Ala Ala Pro Leu Phe Ala Ala Trp His Thr Gly Asp Ala Pro
85 90 95

Arg Val Ala Gly Arg Leu Ala Glu Gly Cys His Ser Met Val Gln Cys
100 105 110

Ala Pro Ala Val Arg Ala Ala Ile Gly Glu Glu Phe Gly Met Pro Ala
115 120 125

Gly Ala Leu Thr Pro Gly Arg Leu Ala Ala Ala Leu Arg Arg Leu Gly
130 135 140

Phe Asp Arg Val Tyr Asp Thr Asn Phe Ala Ala Asp Leu Thr Ile Met
145 150 155 160

Glu Glu Gly Ser Glu Leu Leu Gln Arg Met Glu Gly Ala Gly Pro Leu
165 170 175

Pro Met Phe Thr Ser Cys Cys Pro Ala Trp Val Arg Tyr Ala Glu Gln
180 185 190

Gln Phe Pro Asp Leu Leu Glu His Leu Ser Ser Cys Lys Ser Pro Gln
195 200 205

Gln Met Ala Gly Ala Val Phe Lys Ser Tyr Gly Ala Gln Leu Asp Gly
210 215 220

Val Asp Pro Arg Gln Val Phe Ser Val Ala Val Met Pro Cys Thr Cys
225 230 235 240

Lys Lys Ala Glu Ala Gln Arg Pro Gly Met Glu His Asp Gly Val Arg
245 250 255

Asp Val Asp Ala Val Leu Thr Thr Gly Glu Leu Ala Ala Met Leu Arg
260 265 270

Gln Ala His Ile Asp Phe Ala Ala Leu Pro Asp Glu Pro Phe Asp Arg
275 280 285

Pro Leu Gly Ser Tyr Ser Gly Ala Gly Asn Ile Phe Gly Leu Thr Gly
290 295 300

Gly Val Met Glu Ala Ala Leu Arg Thr Ala Tyr Glu Leu Val Thr Gly
305 310 315 320

Glu Pro Val Pro Cys Thr Glu Leu Val Tyr Val Arg Gly Gly Glu Gly
325 330 335

Ile Arg His Ala Thr Leu Thr Met Asp Gly Arg Thr Phe Arg Val Ala
340 345 350

Val Val Ala Gly Leu Gln His Val Arg Pro Leu Leu Glu Ala Val Arg
355 360 365

Ala Gly Thr Cys Asp Val Asn Phe Val Glu Val Met Cys Cys Pro Gln
370 375 380

Gly Cys Ile Ser Gly Gly Gly Gln Pro Lys Val Leu Leu Pro Phe Gln
385 390 395 400

Arg Asp Glu Val Tyr Ala Ala Arg Lys Ala Ala Leu Tyr Arg His Asp
405 410 415

Ala Glu Leu Ala Cys Arg Lys Ser His Glu Asn Pro Gln Val Gln Ala
420 425 430

Leu Tyr Arg Glu Phe Leu Gly Glu Pro Leu Ser His Val Ser His Asn
435 440 445

Leu Leu His Thr Val Tyr Gly Gln Thr Arg
450 455

<210> 33
<211> 554
<212> PRT
<213> Desulfitobacterium hafniense

<400> 33

Met Met Gln Leu Lys His Pro Phe Gln Ser Gly Phe Gln Gln Gln Ser
1 5 10 15

Cys Lys Arg His Thr Lys Lys Val Val Val Asp Met Glu Ser Lys Ala
20 25 30

Gly Lys Gly Ser Asn Leu Ser Arg Arg Ser Phe Leu Lys Phe Ala Gly
35 40 45

Gly Ala Gly Ile Ala Gly Ala Ser Leu Ser Leu Thr Gly Cys Gly Gln
50 55 60

Pro Leu Thr Pro Ala Ser Ala Val Gly Gly Glu Gly Trp Met Pro Thr
65 70 75 80

Gln Tyr Asn Glu Pro Gly Gly Trp Pro Thr Asn Val Arg Gly Arg Val
85 90 95

Pro Ile Asp Pro Glu Asn Pro Ala Leu Arg Arg Asp Asp Gln Lys Cys
100 105 110

Ile Leu Cys Gly Gln Cys Ile Glu Val Cys Lys Thr Ile Gln Ser Val
115 120 125

Tyr Gly Asn Tyr Glu Leu Pro Leu Lys Asn Glu Ile Pro Cys Ile Asn
130 135 140

Cys Gly Gln Cys Ile His Trp Cys Pro Ser Gly Ala Ile Ser Glu Arg
145 150 155 160

Glu Asp Ile Asp Gln Val Ala Lys Ala Leu Ala Asp Pro Lys Ile Thr
165 170 175

Val Val Val Gln Thr Ala Pro Ala Thr Arg Ile Gly Leu Gly Glu Glu
180 185 190

Phe Gly Leu Pro Val Gly Thr Asn Val Gln Gly Lys Gln Val Ala Ala
195 200 205

Leu Arg Lys Leu Gly Phe Asp Val Ile Phe Asp Thr Asn Phe Ala Ala
210 215 220

Asp Leu Thr Ile Met Glu Glu Gly Thr Glu Leu Val Lys Arg Ile Thr
225 230 235 240

Gly Glu Leu His His Pro Leu Pro Gln Phe Thr Ser Cys Cys Pro Gly
245 250 255

Trp Val Lys Phe Val Glu Tyr Tyr Tyr Pro Glu Leu Leu Pro Asn Leu
260 265 270

Ser Ser Ala Lys Ser Pro Gln Gln Met Ala Gly Ala Leu Val Lys Thr
275 280 285

Tyr Phe Ala Glu Lys Asn His Val Glu Pro Gln Lys Ile Phe Ser Val
290 295 300

Ala Ile Met Pro Cys Thr Ala Lys Lys Phe Glu Cys Gln Arg Pro Glu
305 310 315 320

Met Ile Ser Ala Gln Thr Tyr Trp Gln Asp Glu Gln Val Ser Pro Asp
325 330 335

Val Asp Val Val Leu Thr Thr Arg Glu Leu Ala Arg Met Ile Lys Arg
340 345 350

Ala Gly Ile Asp Leu Pro Ser Leu Pro Asp Glu Glu Tyr Asp Gln Leu
355 360 365

Met Gly Val Ala Thr Gly Ala Gly Ala Ile Phe Gly Thr Thr Gly Gly
370 375 380

Val Met Glu Ala Ala Val Arg Ser Ala Tyr Tyr Leu Val Thr Gly Glu
385 390 395 400

Gln Pro Pro Ala Ala Leu Trp Gln Leu Thr Pro Val Arg Gly Met Glu
405 410 415

Gly Val Lys Glu Ala Ala Val Ser Ile Pro Gly Ala Gly Glu Ile Arg
420 425 430

Ile Ala Val Ile Ser Gly Leu Asp Asn Ala Arg Ala Ile Met Glu Gln
435 440 445

Val Lys Ala Gly Asn Ser Pro Trp Thr Phe Ile Glu Val Met Ala Cys
450 455 460

Pro Gly Gly Cys Gln Tyr Gly Gly Gly Gln Pro Arg Ser Ser Ala Pro
465 470 475 480

Pro Ser Asp Gly Val Arg Asn Thr Arg Ala Ala Ser Leu Tyr Lys Ile
485 490 495

Asp Ala Gln Ala Lys Leu Arg Asn Ser His Asp Asn Pro Gln Ile Lys
500 505 510

Gln Val Tyr Ala Glu Phe Leu Thr Ser Pro Leu Ser Glu Lys Ala Glu
515 520 525

Glu Leu Leu His Thr His Tyr Ile Ser Arg Ala Glu Glu Phe Asp Ala
530 535 540

Lys Lys Pro Gln Ser His Glu Tyr Glu Val
545 550

<210> 34
<211> 578
<212> PRT
<213> Eubacterium acidaminophilum

<400> 34

Met Val Asn Ile Thr Ile Asp Gly Arg Gln Val Thr Val Pro Ala Asn
1 5 10 15

Ser Thr Val Leu Asp Ala Ala Arg Asp Met Gly Ile Asn Ile Pro Thr
20 25 30

Leu Cys Tyr Leu Lys Asp Ile Asn Lys Thr Gly Ala Cys Arg Met Cys
35 40 45

Leu Val Glu Val Glu Gly Ile Arg Asn Leu Gln Thr Ala Cys Thr Phe
50 55 60

Pro Val Arg Asp Gly Leu Val Val Lys Thr Asn Thr Lys Arg Val Arg
65 70 75 80

Asp Ala Arg Arg Asp Asn Leu Gln Leu Ile Leu Ser Asn His His Arg
85 90 95

Asp Cys Leu Ser Cys Phe Arg Asn Gly Ser Cys Glu Leu Gln Ala Leu
100 105 110

Cys Asp Asp Met Gly Leu Ser Glu Leu Asp Phe Glu Ala Pro Lys Glu
115 120 125

Leu Lys Pro Val Asp Met Leu Ser His Ser Ile Val Arg Asp Pro Asn
130 135 140

Lys Cys Ile Leu Cys Gly Arg Cys Val Ala Val Cys Asn Lys Val Gln
145 150 155 160

Glu Val Gly Ile Leu Ala Phe Thr Asn Arg Gly Val Glu Thr Glu Val
165 170 175

Ala Pro Ala Phe Ala Thr Ser Met Ala Asp Ala Pro Cys Ile Tyr Cys
180 185 190

Gly Gln Cys Val Asn Val Cys Pro Val Ala Ala Leu Arg Glu Lys Thr
195 200 205

Asp Ile Glu Lys Val Trp Glu Val Leu Glu Asp Glu Thr Lys His Val
210 215 220

Val Val Gln Val Ala Pro Ala Val Arg Ala Ala Leu Gly Glu Met Phe
225 230 235 240

Gly Asn Pro Ile Gly Thr Arg Val Thr Gly Lys Met Phe Thr Ala Leu
245 250 255

Lys Met Leu Gly Phe Gln Lys Val Phe Asp Thr Asn Phe Ala Ala Asp
260 265 270

Leu Thr Ile Met Glu Glu Gly Thr Glu Leu Leu Gly Arg Ile Lys Asn
275 280 285

Gly Gly Thr Leu Pro Met Ile Thr Ser Cys Ser Pro Gly Trp Ile Arg
290 295 300

Tyr Val Glu His Phe Tyr Pro Glu Leu Leu Asp His Val Ser Ser Cys
305 310 315 320

Lys Ser Pro Gln Gln Met Met Gly Ala Val Leu Lys Ser Tyr Tyr Ala
325 330 335

Glu Lys Asn Asn Ile Ala Pro Glu Asn Met Ile Val Val Ser Val Met
340 345 350

Pro Cys Ile Ala Lys Lys Thr Glu Ser Ala Lys Glu Glu Met Lys Asn
355 360 365

Val His Gly Thr Arg Asp Val Asp Ile Val Leu Thr Thr Arg Glu Leu
370 375 380

Gly Lys Met Ile Lys Glu Ala Arg Ile Glu Phe Asn Asp Leu Gln Asp

385 390 395 400

Ser Asn Pro Asp Glu Phe Phe Gly Asp Tyr Thr Gly Ala Ala Val Ile
405 410 415

Phe Gly Ala Thr Gly Gly Val Met Glu Ala Ala Ile Arg Thr Val Ala
420 425 430

Asp Ile Val Ser Gly Gln Glu Leu Glu Asp Ile Glu Tyr Thr Ala Val
435 440 445

Arg Gly Leu Glu Gly Ile Lys Glu Ala Ala Val Lys Ile Gly Asp Leu
450 455 460

Glu Val Lys Val Ala Val Ala His Gly Thr Ala Asn Ala Gly Lys Leu
465 470 475 480

Met Asp Leu Val Arg Asp Gly Lys Ala Asp Tyr His Phe Ile Glu Ile
485 490 495

Met Gly Cys Ser Gly Gly Cys Val Thr Gly Gly Gly Gln Pro His Val
500 505 510

Asp Ser Arg Thr Lys Glu Lys Val Asn Val Lys Leu Glu Arg Ala Lys
515 520 525

Ala Leu Tyr Thr Glu Asp Lys Leu Arg Asp Lys Arg Lys Ser His His
530 535 540

Asn Glu Ser Val Lys Arg Leu Tyr Glu Glu Tyr Leu Gly Lys Pro Asn
545 550 555 560

Gly His Lys Ala His Glu Leu Leu His Thr His Tyr Lys Lys Arg Glu
565 570 575

Leu Phe

<210> 35
<211> 619
<212> PRT
<213> Rhodopseudomonas palustris

<400> 35

Met Cys Thr Pro Asp Gln Ala Ser Leu Ser Ala Arg Asp Pro Ala Glu
1 5 10 15

Ala Thr Ile Thr Leu Ser Ile Asn Gly Val Ala Cys Ala Gly Phe Ala
20 25 30

Asn Glu Thr Ile Leu Ser Cys Ala Arg Arg Tyr Asp Val Tyr Ile Pro
35 40 45

Thr Leu Cys Glu Leu Glu Asp Ile Asp His Thr Pro Gly Ala Cys Arg
50 55 60

Val Cys Leu Val Glu Ile Leu Gln Ala Gly Lys Asp Thr Pro Gln Ile

Val Thr Ala Cys Asn Thr Pro Val Arg Asp Gly Met Glu Val Gln Thr
85 90 95

Arg Ser Lys Lys Ala Arg Asp Met Gln Arg Leu Gln Val Glu Leu Leu
100 105 110

Met Ala Asp His Leu Gln Asp Cys Ala Thr Cys Ile Arg His Gly Ser
115 120 125

Cys Glu Leu Gln Asp Leu Ala Gln Phe Val Gly Leu Gln Gln Asn Arg
130 135 140

Phe Phe Asp Arg Glu Arg Thr Glu Ala Arg Pro Val Asp His Ser Ser
145 150 155 160

Pro Ser Met Val Arg Asp Met Arg Arg Cys Val Arg Cys Gln Arg Cys
165 170 175

Val Ala Ile Cys Arg Tyr His Gln Lys Ile Asp Ala Leu Ala Ile Glu
180 185 190

Gly Ser Gly Leu Glu Arg Met Val Ala Leu Arg Asp Ala Asp Gly Tyr
195 200 205

Pro Asn Ser Val Cys Val Ser Cys Gly Gln Cys Val Leu Val Cys Pro
210 215 220

Thr Gly Ala Leu Gly Glu Arg Asp Glu Thr Asp Arg Ala Leu Asp Tyr
225 230 235 240

Ile Cys Asp Pro Asn Val Val Thr Val Val Gln Phe Ala Pro Ala Val
245 250 255

Arg Val Ala Phe Gly Glu Glu Phe Gly Leu Pro Ala Gly Thr Asn Val
260 265 270

Glu Gly Gln Ile Ile Ala Ala Cys Arg Lys Leu Gly Val Asp Val Val
275 280 285

Leu Asp Thr Asn Phe Ala Ala Asp Val Val Ile Met Glu Glu Gly Ala
290 295 300

Glu Leu Leu Ala Arg Leu Lys Gln Gly Arg Arg Pro Thr Phe Thr Ser
305 310 315 320

Cys Cys Pro Ala Trp Ile Asn Phe Ala Glu Ile His Tyr Pro Asp Val
325 330 335

Leu Pro Leu Leu Ser Ser Thr Lys Ser Pro Gln Gln Val Leu Ser Thr
340 345 350

Ile Ala Lys Ser Tyr Leu Pro Ala Gln Leu Gly Val Pro Ala Glu Arg
355 360 365

Ile Arg Val Ile Ser Ile Met Pro Cys Ile Ala Lys Lys Asp Glu Ala
370 375 380

Val Arg Pro Gln Met Val His Asp Gly Gln Pro Glu Thr Asp Leu Val
385 390 395 400

Leu Thr Thr Arg Glu Phe Ala Arg Leu Leu Arg Arg Glu Gly Ile Asp
405 410 415

Leu Lys Asp Leu Pro Ser Ser Gln Phe Asp Arg Pro Phe Leu Ser Ala
420 425 430

Tyr Ser Gly Ala Gly Ala Ile Phe Gly Thr Thr Gly Gly Val Met Glu
435 440 445

Ala Ala Val Arg Thr Ile Tyr Ala Leu Val Asn Gly Arg Glu Leu Glu
450 455 460

Arg Ile Glu Leu Thr Gln Leu Arg Gly Phe Glu Gly Leu Arg Glu Ala
465 470 475 480

Thr Val Asp Leu Gly Ala Pro Val Gly Glu Val Lys Val Ala Met Val
485 490 495

His Gly Leu Gly Asp Thr Arg Lys Leu Val Glu Ser Val Leu Ser Gly
500 505 510

Glu Ala Asn Tyr Asp Phe Ile Glu Val Met Ala Cys Pro Gly Gly Cys
515 520 525

Val Asp Gly Gly Gly Ser Leu Arg Ser Lys Lys Ala Tyr Leu Pro Leu
530 535 540

Ala Leu Lys Arg Arg Glu Thr Ile Tyr Asn Val Asp Arg Ala Ala Lys
545 550 555 560

Val Arg Gln Ser His Asn Asn Pro Gln Val Gln Ala Leu Tyr Arg Glu
565 570 575

Leu Leu Gln Ala Pro Asn Ser Glu Ile Ala His Arg Leu Leu His Thr
580 585 590

His Tyr Ala Ser Arg Lys Arg Glu Leu Gln His Thr Val Lys Glu Ile
595 600 605

Trp Asp Asp Leu Thr Met Ser Thr Ile Leu Tyr
610 615

<210> 36
<211> 644
<212> PRT
<213> Clostridium thermocellum

<400> 36

Met Asp Ser Phe Leu Met Lys Gly Tyr Ile Lys Glu Ala Asn Ile Asp
1 5 10 15

Tyr Ser Cys Ser Arg Gly Ser Met Glu Asp Leu Pro Lys Trp Glu Phe
20 25 30

Arg Glu Ile Pro Lys Val Pro Arg Ala Val Met Pro Ser Leu Ser Leu
35 40 45

Glu Glu Arg Lys Asn Asn Phe Asn Glu Val Glu Leu Gly Leu Ser Glu
50 55 60

Glu Val Ala Arg Lys Glu Ala Arg Arg Cys Leu Lys Cys Gly Cys Ser
65 70 75 80

Ala Arg Phe Thr Cys Asp Leu Arg Lys Glu Ala Ser Asn His Gly Ile
85 90 95

Val Tyr Glu Glu Pro Ile His Asp Arg Pro Tyr Ile Pro Lys Val Asp
100 105 110

Asp His Pro Phe Ile Val Arg Asp His Asn Lys Cys Ile Ser Cys Gly
115 120 125

Arg Cys Ile Ala Ala Cys Ala Glu Ile Glu Gly Pro Gly Val Leu Thr
130 135 140

Phe Tyr Met Lys Asn Gly Arg Gln Leu Val Gly Thr Lys Ser Gly Leu
145 150 155 160

Pro Leu Arg Asp Thr Asp Cys Val Ser Cys Gly Gln Cys Val Thr Ala
165 170 175

Cys Pro Cys Ala Ala Leu Asp Tyr Arg Arg Glu Arg Gly Lys Val Val
180 185 190

Arg Ala Ile Asn Asp Pro Lys Lys Thr Val Val Gly Phe Val Ala Pro
195 200 205

Ala Val Arg Ser Leu Ile Ser Asn Thr Phe Gly Val Ser Tyr Glu Glu
210 215 220

Ala Ser Pro Phe Met Ala Gly Leu Leu Lys Lys Leu Gly Phe Asp Lys
225 230 235 240

Val Phe Asp Phe Thr Phe Ala Ala Asp Leu Thr Ile Val Glu Glu Thr
245 250 255

Thr Glu Phe Leu Ser Arg Ile Gln Asn Lys Gly Val Met Pro Gln Phe
260 265 270

Thr Ser Cys Cys Pro Gly Trp Ile Asn Phe Val Glu Lys Arg Tyr Pro
275 280 285

Glu Ile Ile Pro His Leu Ser Thr Cys Lys Ser Pro Gln Met Met Met
290 295 300

Gly Ala Thr Val Lys Asn His Tyr Ala Lys Leu Met Gly Ile Asn Lys
305 310 315 320

WO 2005/072262 PCT/US2005/001983

Glu Asp Leu Phe Val Val Ser Ile Val Pro Cys Leu Ala Lys Lys Tyr
325 330 335

Glu Ala Ala Arg Pro Glu Phe Ile His Asp Gly Ile Arg Asp Val Asp
340 345 350

Ala Val Leu Thr Thr Thr Glu Met Leu Glu Met Met Glu Leu Ala Asp
355 360 365

Ile Lys Pro Ser Glu Val Val Pro Gln Glu Phe Asp Glu Pro Tyr Lys
370 375 380

Gln Val Ser Gly Ala Gly Ile Leu Phe Gly Ala Ser Gly Gly Val Ala
385 390 395 400

Glu Ala Ala Leu Arg Met Ala Val Glu Lys Leu Thr Gly Lys Val Leu
405 410 415

Thr Asp His Leu Glu Phe Glu Glu Ile Arg Gly Phe Glu Gly Val Lys
420 425 430

Glu Ser Thr Ile Asp Val Asn Gly Thr Lys Val Arg Val Ala Val Val
435 440 445

Ser Gly Leu Lys Asn Ala Glu Pro Ile Ile Glu Lys Ile Leu Asn Gly
450 455 460

Val Asp Val Gly Tyr Asp Leu Ile Glu Val Met Ala Cys Pro Gly Gly
465 470 475 480

Cys Ile Cys Gly Ala Gly His Pro Val Pro Glu Lys Ile Asp Ser Leu
485 490 495

Glu Lys Arg Gln Gln Val Leu Val Asn Ile Asp Lys Val Ser Lys Tyr
500 505 510

Arg Lys Ser Gln Glu Asn Pro Asp Ile Leu Arg Leu Tyr Asn Glu Phe
515 520 525

Tyr Gly Glu Pro Asn Ser Pro Leu Ala His Glu Leu Leu His Thr His
530 535 540

Tyr Thr Pro Lys His Gly Asp Ser Thr Cys Ser Pro Glu Arg Lys Lys
545 550 555 560

Gly Thr Ala Ala Phe Asp Val Gln Glu Phe Thr Ile Cys Met Cys Glu
565 570 575

Ser Cys Met Glu Lys Gly Ala Glu Asn Leu Tyr Asn Asp Leu Ser Ser
580 585 590

Lys Ile Arg Leu Phe Lys Met Asp Pro Phe Val Gln Ile Lys Arg Ile
595 600 605

Arg Leu Lys Glu Thr His Pro Gly Lys Gly Val Tyr Ile Ala Leu Asn
610 615 620

Gly Lys Gln Ile Glu Glu Pro Met Leu Ser Gly Asn Ile Pro Asp Glu
625 630 635 640

Ser Glu Ser Glu

<210> 37
<211> 572
<212> PRT
<213> Clostridium perfringens

<400> 37

Met Asn Lys Ile Ile Ile Asn Asp Lys Thr Ile Glu Phe Asp Gly Asp
1 5 10 15

Lys Thr Ile Leu Asp Leu Ala Arg Glu Asn Gly Phe Asp Ile Pro Val
20 25 30

Leu Cys Glu Leu Lys Asn Cys Gly Asn Lys Gly Gln Cys Gly Val Cys
35 40 45

Leu Val Glu Gln Glu Gly Asn Asp Arg Leu Leu Arg Ser Cys Ala Ile
50 55 60

Lys Ala Lys Asp Gly Met Val Ile Lys Thr Asp Ser Glu Lys Val Leu
65 70 75 80

Glu Ala Arg Lys Glu Arg Val Ala Glu Leu Leu Asp Glu His Glu Phe
85 90 95

Lys Cys Gly Pro Cys Lys Arg Arg Glu Asn Cys Glu Phe Leu Lys Leu
100 105 110

Val Ile Lys Thr Lys Ala Arg Ala His Lys Pro Phe Val Val Ala Asp
115 120 125

Lys Ser Glu Tyr Val Asp Asp Arg Ser Lys Ser Ile Val Leu Asp Arg
130 135 140

Ser Lys Cys Val Lys Cys Gly Arg Cys Val Ala Ala Cys Arg Thr Arg
145 150 155 160

Thr Ala Thr Asn Ser Ile Lys Phe His Arg Ile Asp Gly Val Arg Leu
165 170 175

Val Gly Pro Glu Glu Leu Lys Cys Phe Asp Asp Thr Asn Cys Leu Leu
180 185 190

Cys Gly Gln Cys Ile Ala Ala Cys Pro Val Asp Ala Leu Ser Glu Lys
195 200 205

Ser His Ile Glu Arg Val Gln Glu Ala Leu Asn Asp Pro Glu Lys His
210 215 220

Val Ile Val Ala Met Ala Pro Ala Val Arg Thr Ser Met Gly Glu Leu
225 230 235 240

Phe Lys Met Gly Tyr Gly Gln Asp Val Thr Gly Lys Leu Tyr Thr Ala
245 250 255

Leu Arg Glu Leu Gly Phe Asp Lys Val Phe Asp Ile Asn Phe Gly Ala
260 265 270

Asp Met Thr Ile Met Glu Glu Ala Thr Glu Leu Ile Glu Arg Ile Lys
275 280 285

Asn Asn Gly Pro Phe Pro Met Leu Thr Ser Cys Cys Pro Ser Trp Val
290 295 300

Arg Glu Val Glu Asn Tyr Phe Pro Glu Leu Val Glu Asn Leu Ser Ser
305 310 315 320

Ala Lys Ser Pro Gln Gln Ile Phe Gly Ala Ala Ser Lys Thr Tyr Tyr
325 330 335

Pro Gln Val Ala Asp Ile Asp Pro Lys Lys Val Phe Thr Val Thr Val
340 345 350

Met Pro Cys Thr Ser Lys Lys Phe Glu Ala Asp Arg Pro Glu Met Glu
355 360 365

Asn Glu Gly Ile Arg Asn Ile Asp Ala Val Ile Thr Thr Arg Glu Leu
370 375 380

Ala Arg Met Ile Lys Ala Ala Lys Ile Asp Phe Ala Lys Leu Glu Asp
385 390 395 400

Gly Glu Val Asp Pro Ala Met Gly Glu Tyr Thr Gly Ala Gly Val Ile
405 410 415

Phe Gly Ala Thr Gly Gly Val Met Glu Ala Ala Leu Arg Thr Ala Lys
420 425 430

Asp Phe Met Glu Asn Asp Asn Leu Asp Asn Val Asp Tyr Glu Ala Val
435 440 445

Arg Gly Leu Ala Gly Ile Lys Glu Ala Glu Val Glu Ile Ala Gly Asn
450 455 460

Glu Tyr Lys Leu Ala Val Val Ser Gly Ala Ala Asn Val Phe Glu Leu
465 470 475 480

Val Lys Ser Gly Lys Ile Asn Asp Tyr His Phe Ile Glu Val Met Ala
485 490 495

Cys Pro Gly Gly Cys Val Asn Gly Gly Gly Gln Pro His Ile Ser Ala
500 505 510

Glu Asp Ser Asp Lys Met Asp Ile Arg Glu Val Arg Ala Ser Val Leu
515 520 525

Tyr Asn Gln Asp Lys Asn Leu Glu Lys Arg Lys Ser His Gln Asn Ser

530 535 540

Ala Leu Leu Lys Met Tyr Glu Ser Tyr Met Gly Lys Pro Gly His Gly
545 550 555 560

Arg Ala His Glu Leu Leu His Met Lys Tyr Lys Lys
565 570

<210> 38
<211> 583
<212> PRT
<213> Clostridium thermocellum

<400> 38

Met His Val Leu Lys Leu Val His Ser Thr Gln Tyr Trp Arg Ala Glu
1 5 10 15

Glu Met Asp Asn Arg Glu Tyr Met Leu Ile Asp Gly Ile Pro Val Glu
20 25 30

Ile Asn Gly Glu Lys Asn Leu Leu Glu Leu Ile Arg Lys Ala Gly Ile
35 40 45

Lys Leu Pro Thr Phe Cys Tyr His Ser Glu Leu Ser Val Tyr Gly Ala
50 55 60

Cys Arg Met Cys Met Val Glu Asn Glu Trp Gly Gly Leu Asp Ala Ala
65 70 75 80

Cys Ser Thr Pro Pro Arg Ala Gly Met Ser Ile Lys Thr Asn Thr Glu
85 90 95

Arg Leu Gln Lys Tyr Arg Lys Met Ile Leu Glu Leu Leu Leu Ala Asn
100 105 110

His Cys Arg Asp Cys Thr Thr Cys Asn Asn Asn Gly Lys Cys Lys Leu
115 120 125

Gln Asp Leu Ala Met Arg Tyr Asn Ile Ser His Ile Arg Phe Pro Asn
130 135 140

Thr Ala Ser Asn Pro Asp Val Asp Asp Ser Ser Leu Cys Ile Thr Arg
145 150 155 160

Asp Arg Ser Lys Cys Ile Leu Cys Gly Asp Cys Val Arg Val Cys Asn
165 170 175

Glu Val Gln Asn Val Gly Ala Ile Asp Phe Ala Tyr Arg Gly Ser Lys
180 185 190

Met Thr Ile Ser Thr Val Phe Asp Lys Pro Ile Phe Glu Ser Asn Cys
195 200 205

Val Gly Cys Gly Gln Cys Ala Leu Ala Cys Pro Thr Gly Ala Ile Val
210 215 220

Val Lys Asp Asp Thr Gln Lys Val Trp Lys Glu Ile Tyr Asp Lys Asn

225 230 235 240

Thr Arg Val Ser Val Gln Ile Ala Pro Ala Val Arg Val Ala Leu Gly
245 250 255

Lys Glu Leu Gly Leu Asn Asp Gly Glu Asn Ala Ile Gly Lys Ile Val
260 265 270

Ala Ala Leu Arg Arg Met Gly Phe Asp Asp Ile Phe Asp Thr Ser Thr
275 280 285

Gly Ala Asp Leu Thr Val Leu Glu Glu Ser Ala Glu Leu Leu Arg Arg
290 295 300

Ile Arg Glu Gly Lys Asn Asp Met Pro Leu Phe Thr Ser Cys Cys Pro
305 310 315 320

Ala Trp Val Asn Tyr Cys Glu Lys Phe Tyr Pro Glu Leu Leu Pro His
325 330 335

Val Ser Thr Cys Arg Ser Pro Met Gln Met Phe Ala Ser Ile Ile Lys
340 345 350

Glu Glu Tyr Ser Thr Ser Ser Lys Arg Leu Val His Val Ala Val Met
355 360 365

Pro Cys Thr Ala Lys Lys Phe Glu Ala Ala Arg Lys Glu Phe Lys Val
370 375 380

Asn Gly Val Pro Asn Val Asp Tyr Val Leu Thr Thr Gln Glu Leu Val
385 390 395 400

Arg Met Ile Lys Glu Ser Gly Ile Val Phe Ser Glu Leu Glu Pro Glu
405 410 415

Ala Ile Asp Met Pro Phe Gly Thr Tyr Thr Gly Ala Gly Val Ile Phe
420 425 430

Gly Val Ser Gly Gly Val Thr Glu Ala Val Leu Arg Arg Val Val Ser
435 440 445

Asp Lys Ser Pro Thr Ser Phe Arg Ser Leu Ala Tyr Thr Gly Val Arg
450 455 460

Gly Met Asn Gly Val Lys Glu Ala Ser Val Met Tyr Gly Asp Arg Lys
465 470 475 480

Leu Lys Val Ala Val Val Ser Gly Leu Lys Asn Ala Gly Asp Leu Ile
485 490 495

Glu Arg Ile Lys Ala Gly Glu His Tyr Asp Leu Val Glu Val Met Ala
500 505 510

Cys Pro Gly Gly Cys Ile Asn Gly Gly Gly Gln Pro Phe Val Gln Ser
515 520 525

Glu Glu Arg Glu Lys Arg Gly Lys Gly Leu Tyr Ser Ala Asp Lys Leu
530 535 540

Cys Asn Ile Lys Ser Ser Glu Glu Asn Pro Leu Met Met Thr Leu Tyr
545 550 555 560

Lys Gly Ile Leu Lys Gly Arg Val His Glu Leu Leu His Val Asp Tyr
565 570 575

Ala Ser Lys Lys Glu Ala Lys
580

<210> 39
<211> 439
<212> PRT
<213> Desulfovibrio desulfuricans

<400> 39

Met Ala Gly Cys Lys Ala Gln His Pro Pro Ala Ala Tyr Leu Ala Gly
1 5 10 15

Leu Glu Val Pro Ala Ala Gly Ser Glu Val Thr Met Glu Gly Val Arg
20 25 30

Tyr Lys Met Asn Ala Pro Lys Asp Val Asp Pro Ala Thr Ile Arg Phe
35 40 45

Val Glu Val Asp His Asp Lys Cys Met Ala Cys Gly Glu Cys Glu Tyr
50 55 60

His Cys Pro Thr Gly Val Met Gln Glu Val Thr Glu Asp Gly Tyr Arg
65 70 75 80

Gly Val Val Asp Pro Val Ala Cys Val Asn Cys Gly Gln Cys Leu Ala
85 90 95

Asn Cys Pro Phe Gly Ala Ile His Glu Glu Val Ser Phe Val Gly Glu
100 105 110

Leu Tyr Glu Lys Leu Lys Asp Pro Asp Thr Val Val Val Ser Met Pro
115 120 125

Ala Pro Ala Val Arg Tyr Ala Leu Gly Glu Cys Phe Gly Leu Pro Thr
130 135 140

Gly Thr Tyr Val Gly Gly Gln Met His Ala Ala Leu Arg Arg Leu Gly
145 150 155 160

Phe Asn Leu Val Trp Asp Thr Glu Trp Thr Ala Asp Val Thr Ile Met
165 170 175

Glu Glu Gly Thr Glu Leu Leu Glu Arg Val Lys His Gly Asn Met Pro
180 185 190

Leu Pro Gln Phe Thr Ser Cys Cys Pro Gly Trp Ile Lys Phe Ala Glu
195 200 205

Thr Phe Tyr Pro Asp Leu Glu Lys His Leu Ser Thr Cys Lys Ser Pro
210 215 220

Ile Ala Met Ile Gly Pro Leu Ala Lys Thr Tyr Gly Ala Gln Glu Ala
225 230 235 240

Gly Val Pro Ala Lys Lys Met Tyr Thr Val Ser Ile Met Pro Cys Ile
245 250 255

Ala Lys Lys Phe Glu Gly Met Arg Pro Glu Met Asn Ala Ser Gly Tyr
260 265 270

Arg Asp Ile Asp Ala Thr Ile Thr Thr Arg Glu Leu Ala Trp Met Ile
275 280 285

Lys Lys Ala Gly Ile Asp Phe Thr Ser Leu Pro Ser Glu Glu Pro Asp
290 295 300

Pro Ala Leu Gly Met Ser Thr Gly Ala Ala Thr Ile Phe Cys Thr Ser
305 310 315 320

Gly Gly Val Met Glu Ala Ala Leu Arg Leu Ala Tyr Glu Ala Leu Ser
325 330 335

Gly Gly Thr Leu Ala Asp Pro Asp Ile Lys Val Val Arg Thr His Glu
340 345 350

Gly Ile Asn Thr Ala Glu Val Pro Val Pro Asn Phe Gly Thr Val Lys
355 360 365

Val Ala Val Ala Ser Gly Leu Asp Asn Ala Ala Lys Leu Cys Glu Glu
370 375 380

Val Arg Ala Gly Lys Ser Pro Tyr His Phe Ile Glu Val Met Thr Cys
385 390 395 400

Pro Gly Gly Cys Val Asn Gly Gly Gly Gln Pro Leu Glu Pro Gly Met
405 410 415

Leu Gln Ser Ser Leu Phe Lys Ser Thr Ile Thr Lys Ile Asn Arg Arg
420 425 430

Phe Thr Arg Arg Ser Val Ala
435

<210> 40
<211> 379
<212> PRT
<213> Desulfovibrio desulfuricans

<400> 40

Met Asn Leu Val Glu Met Glu Lys Ile Gln Tyr Val Asp Gln Ser Pro
1 5 10 15

Asp Pro Arg Ala Asn Pro Asp Glu Leu Phe Phe Ile Gln Ile Asp Pro
20 25 30

Glu Lys Cys Ile Gly Cys Asp Thr Cys Gln Glu Tyr Cys Pro Thr Gly
35 40 45

Ala Ile Phe Gly Asp Thr Gly Ser Ala His Ser Ile Pro His Glu Glu
50 55 60

Ile Cys Ile Asn Cys Gly Gln Cys Leu Thr His Cys Pro Val Gly Ala
65 70 75 80

Ile Tyr Glu Val Gln Ser Trp Val Arg Glu Leu Ser Glu Lys Ile Lys
85 90 95

Asp Pro Glu Ile Lys Val Ile Ala Met Pro Ala Pro Ala Val Arg Tyr
100 105 110

Gly Leu Gly Glu Cys Phe Gly Met Pro Val Gly Thr Val Thr Thr Gly
115 120 125

Lys Met Leu Thr Ala Leu Gln Met Leu Gly Phe Asp His Val Trp Asp
130 135 140

Asn Glu Phe Thr Ala Asp Val Thr Ile Trp Glu Glu Gly Thr Glu Phe
145 150 155 160

Val Lys Arg Leu Thr Gly Gln Ile Asp Lys Pro Leu Pro Gln Phe Thr
165 170 175

Ser Cys Cys Pro Gly Trp His Lys Tyr Val Glu Ser Phe Tyr Pro Glu
180 185 190

Leu Phe Pro His Leu Ser Ser Cys Lys Ser Pro Ile Gly Met Met Gly
195 200 205

Ala Leu Ala Lys Thr Tyr Gly Pro Asp Val Met Lys Tyr Asp Arg Ser
210 215 220

Lys Val Tyr Thr Val Ser Ile Met Pro Cys Thr Ala Lys Lys Tyr Glu
225 230 235 240

Gly Met Arg Ala Asp Leu Trp Ser Ser Gly Tyr Lys Asp Ile Asp Ala
245 250 255

Thr Ile Asp Thr Arg Glu Leu Ala Tyr Met Ile Lys Lys Ala Gly Ile
260 265 270

Asp Phe Ala Ala Leu Pro Asp Gly Lys Arg Asp Thr Leu Met Gly Asp
275 280 285

Ser Thr Gly Gly Ala Thr Ile Phe Gly Val Ser Gly Gly Val Met Glu
290 295 300

Ala Ala Leu Arg Tyr Ala Tyr Glu Ala Val Thr Gly Lys Lys Pro Ser
305 310 315 320

Ser Trp Asp Phe Thr Met Val Arg Gly Leu Asn Gly Ile Lys Glu Gly
325 330 335

Thr Val Thr Ile Gly Asp Ala Lys Ile Asn Val Ala Val Val His Gly
340 345 350

Ala Lys Arg Phe Ala Glu Val Cys Glu Val Ile Lys Thr Gly Lys Ser
355 360 365

Pro Cys Ile Ser Ser Ser Leu Cys Leu Pro Arg
370 375

<210> 41
<211> 421
<212> PRT
<213> Desulfovibrio desulfuricans

<400> 41

Met Asn Leu Val Glu Met Glu Lys Ile Gln Tyr Val Asp Gln Ser Pro
1 5 10 15

Asp Pro Arg Ala Asn Pro Asp Glu Leu Phe Phe Ile Gln Ile Asp Pro
20 25 30

Glu Lys Cys Ile Gly Cys Asp Thr Cys Gln Glu Tyr Cys Pro Thr Gly
35 40 45

Ala Ile Phe Gly Asp Thr Gly Ser Ala His Ser Ile Pro His Glu Glu
50 55 60

Ile Cys Ile Asn Cys Gly Gln Cys Leu Thr His Cys Pro Val Gly Ala
65 70 75 80

Ile Tyr Glu Val Gln Ser Trp Val Arg Glu Leu Ser Glu Lys Ile Lys
85 90 95

Asp Pro Glu Ile Lys Val Ile Ala Met Pro Ala Pro Ala Val Arg Tyr
100 105 110

Gly Leu Gly Glu Cys Phe Gly Met Pro Val Gly Thr Val Thr Thr Gly
115 120 125

Lys Met Leu Thr Ala Leu Gln Met Leu Gly Phe Asp His Val Trp Asp
130 135 140

Asn Glu Phe Thr Ala Asp Val Thr Ile Trp Glu Glu Gly Thr Glu Phe
145 150 155 160

Val Lys Arg Leu Thr Gly Gln Ile Asp Lys Pro Leu Pro Gln Phe Thr
165 170 175

Ser Cys Cys Pro Gly Trp His Lys Tyr Val Glu Ser Phe Tyr Pro Glu
180 185 190

Leu Phe Pro His Leu Ser Ser Cys Lys Ser Pro Ile Gly Met Met Gly
195 200 205

Ala Leu Ala Lys Thr Tyr Gly Pro Asp Val Met Lys Tyr Asp Arg Ser
210 215 220

Lys Val Tyr Thr Val Ser Ile Met Pro Cys Thr Ala Lys Lys Tyr Glu
225 230 235 240

Gly Met Arg Ala Asp Leu Trp Ser Ser Gly Tyr Lys Asp Ile Asp Ala
245 250 255

Thr Ile Asp Thr Arg Glu Leu Ala Tyr Met Ile Lys Lys Ala Gly Ile
260 265 270

Asp Phe Ala Ala Leu Pro Asp Gly Lys Arg Asp Thr Leu Met Gly Asp
275 280 285

Ser Thr Gly Gly Ala Thr Ile Phe Gly Val Ser Gly Gly Val Met Glu
290 295 300

Ala Ala Leu Arg Tyr Ala Tyr Glu Ala Val Thr Gly Lys Lys Pro Ser
305 310 315 320

Ser Trp Asp Phe Thr Met Val Arg Gly Leu Asn Gly Ile Lys Glu Gly
325 330 335

Thr Val Thr Ile Gly Asp Ala Lys Ile Asn Val Ala Val Val His Gly
340 345 350

Ala Lys Arg Phe Ala Glu Val Cys Glu Val Ile Lys Thr Gly Lys Ser
355 360 365

Pro Trp His Phe Ile Glu Phe Met Ala Cys Pro Gly Gly Cys Val Cys
370 375 380

Gly Gly Gly Gln Pro Val Met Pro Gly Val Leu Glu Ala Met Asp Arg
385 390 395 400

Lys Val Ser Arg Thr Phe Ala Gly Leu Lys Glu Arg Leu Asn Arg Met
405 410 415

Ser Ser Ser Lys Ala
420

<210> 42
<211> 369
<212> PRT
<213> Trichomonas vaginalis

<400> 42

Cys Asp Gly Lys Trp Leu Ser Pro Ala Cys Val Thr Thr Val Trp Asp
1 5 10 15

Gly Leu Lys Ile Asp Thr Lys Ser Lys Asn Val Arg Asp Ser Val Glu
20 25 30

Asn Asn Leu Lys Glu Leu Leu Asp Cys His Asp Glu Thr Cys Ser Ala
35 40 45

Cys Ile Ala Asn His Arg Cys Gln Phe Arg Asp Met Asn Val Ala Tyr
50 55 60

Ser Val Lys Ala Glu Thr Lys Glu Ile Cys Ser Glu Glu Gly Ile Asp
65 70 75 80

Glu Ser Thr Asn Ala Ile Arg Leu Asp Thr Ser Lys Cys Val Leu Cys
85 90 95

Gly Arg Cys Ile Arg Ala Cys Glu Glu Val Ala Gly Thr Ser Ala Ile
100 105 110

Ile Phe Gly Asn Arg Ala Lys Lys Met Arg Ile Gln Pro Thr Phe Gly
115 120 125

Val Thr Leu Gln Glu Thr Ser Cys Ile Lys Cys Gly Gln Cys Thr Leu
130 135 140

Tyr Cys Pro Val Gly Ala Ile Thr Glu Lys Ser Gln Val Lys Glu Ala
145 150 155 160

Leu Asp Ile Leu Ala Asn Lys Gly Lys Lys Ile Thr Val Val Gln Val
165 170 175

Ala Pro Ala Val Arg Val Ala Leu Ser Glu Ala Phe Gly Tyr Lys Glu
180 185 190

Gly Thr Val Thr Thr Gly Lys Met Val Ser Ala Leu Lys Ala Leu Gly
195 200 205

Phe Asp Leu Val Tyr Asp Thr Asn Tyr Gly Ala Asp Leu Thr Ile Cys
210 215 220

Glu Glu Ala Gly Glu Leu Val Asn Arg Leu Arg Asp Pro Asn Ala Lys
225 230 235 240

Phe Pro Met Phe Thr Ser Cys Cys Pro Ala Trp Val Asn Tyr Val Glu
245 250 255

Gln Ser Ala Pro Asp Phe Ile Pro Asn Leu Ser Ser Cys Arg Ser Pro
260 265 270

Gln Gly Met Leu Ser Ala Leu Ile Lys Asn Tyr Leu Pro Lys Leu Leu
275 280 285

Asp Val Lys Gln Glu Asp Val Leu Asn Phe Ser Ile Met Pro Cys Thr
290 295 300

Ala Lys Lys Asp Glu Val Glu Arg Pro Glu Leu Arg Thr Lys Ser Gly
305 310 315 320

Pro Lys Glu Thr Asp Met Val Leu Thr Val Arg Glu Leu Val Glu Met
325 330 335

Ile Lys Leu Ser Asn Ile Asp Phe Asn Asn Leu Pro Asp Thr Gln Phe
340 345 350

Asp Asn Ile Phe Gly Phe Gly Ser Gly Ala Gly Gln Ile Phe Ala Ala
355 360 365

Thr

<210> 43
<211> 369
<212> PRT
<213> Trichomonas gallinae

<400> 43

Cys Asp Gly Lys Trp Leu Ser Pro Ala Cys Val Thr Thr Val Trp Asp
1 5 10 15

Gly Leu Arg Ile Asp Thr Lys Ser Lys Val Val Arg Asp Ser Val Glu
20 25 30

Asn Asn Leu Lys Glu Leu Leu Asp Cys His Asp Glu Thr Cys Ser Ser
35 40 45

Cys Val Ala Asn His Arg Cys Gln Phe Arg Asp Met Asn Val Ala Tyr
50 55 60

Ser Val Lys Ala Asp Thr Lys Glu Ile Cys Ser Glu Glu Gly Ile Asp
65 70 75 80

Glu Ser Thr His Ala Ile Arg Leu Asp Thr Ser Lys Cys Val Leu Cys
85 90 95

Gly Arg Cys Ile Arg Ala Cys Glu Glu Val Ala Gly Thr Ser Ala Ile
100 105 110

Ile Phe Gly Asn Arg Ala Lys His Met Arg Ile Gln Pro Thr Phe Gly
115 120 125

Gly Thr Leu Gln Glu Thr Ala Cys Ile Lys Cys Gly Gln Cys Thr Leu
130 135 140

Tyr Cys Pro Val Gly Ala Ile Thr Glu Lys Ser Gln Val Lys Glu Ala
145 150 155 160

Leu Asp Ile Leu Ala Asn Lys Gly Lys Lys Val Thr Val Val Gln Val
165 170 175

Ala Pro Ala Val Arg Val Ala Leu Ser Glu Ala Phe Gly Tyr Lys Glu
180 185 190

Gly Thr Val Thr Thr Gly Lys Met Val Ser Ala Leu Lys Ala Leu Gly
195 200 205

Phe Asp Leu Val Tyr Asp Thr Asn Tyr Gly Ala Asp Leu Thr Ile Cys
210 215 220

Glu Glu Ala Gly Glu Leu Val Asn Arg Leu Lys Asp Pro Lys Ala Val
225 230 235 240

Phe Pro Met Phe Thr Ser Cys Cys Pro Ala Trp Val Asn Tyr Val Glu
245 250 255

Gln Ser Ala Pro Asp Phe Ile Pro Asn Leu Ser Ser Cys Arg Ser Pro
260 265 270

Gln Gly Met Leu Ser Ser Leu Ile Lys Asn Tyr Leu Pro Lys Leu Leu
275 280 285

Gly Ile Lys Gln Glu Glu Val Met Asn Phe Ser Ile Met Pro Cys Thr
290 295 300

Ala Lys Lys Asp Glu Ile Glu Arg Pro Glu Leu Gln Thr Lys Thr Gly
305 310 315 320

Leu Lys Glu Thr Asp Met Val Leu Thr Val Arg Glu Leu Val Glu Met
325 330 335

Ile Lys Leu Ser Asn Ile Asp Phe Asn Asn Leu Pro Asp Thr Pro Phe
340 345 350

Asp Asn Ile Phe Gly Phe Gly Ser Gly Ala Gly Gln Ile Phe Ala Ala
355 360 365

Thr

<210> 44
<211> 456
<212> PRT
<213> Nyctotherus ovalis

<400> 44

Met Ile Ser Arg Leu Ile Ala Lys Lys Ala Pro Leu Phe Leu Arg Thr
1 5 10 15

Phe Ala Thr Ser Glu Met Ile Ser Leu Lys Ile Asp Gly Lys Ile Ile
20 25 30

Ser Val Pro Lys Gly Ile Met Leu Ala Asp Ala Ile Lys Lys Ala Gly
35 40 45

Ala Asn Val Pro Thr Met Cys Tyr His Pro Asp Leu Pro Thr Ser Gly
50 55 60

Gly Ile Cys Arg Val Cys Leu Val Glu Ser Ala Lys Ser Pro Gly Tyr
65 70 75 80

Pro Ile Ile Ser Cys Arg Thr Pro Val Glu Glu Gly Met Glu Ile Val
85 90 95

Thr Gln Gly Ser Lys Met Lys Glu Tyr Arg Gln Ala Asn Leu Ala Leu
100 105 110

Met Leu Ser Arg His Pro Asn Ala Cys Leu Ser Cys Thr Ser Asn Thr
115 120 125

Asn Cys Lys Thr Gln Glu Leu Ser Ala Asn Met Asn Ile Gly Gln Cys
130 135 140

Gly Phe Ala Asn Ala Thr Pro Pro Lys Asn Asp Asp Ser Tyr Asp Met
145 150 155 160

Thr Thr Ala Ile Glu Arg Asp Asn Asp Lys Cys Ile Asn Cys Asp Ile
165 170 175

Cys Val His Thr Cys Ser Leu Gln Gly Leu Asn Ala Leu Gly Phe Tyr
180 185 190

Asn Glu Glu Gly His Ala Val Lys Ser Met Gly Thr Leu Asp Val Ser
195 200 205

Glu Cys Ile Gln Cys Gly Gln Cys Ile Asn Arg Cys Pro Thr Gly Ala
210 215 220

Ile Thr Glu Lys Ser Glu Ile Arg Pro Val Leu Asp Ala Ile Asn Ile
225 230 235 240

Gln Gln Arg Leu Val Phe Gln Met Ala Pro Ser Ile Arg Val Ala Val
245 250 255

Ala Glu Glu Phe Gly Ile Lys Pro Gly Glu Lys Ile Leu Lys Asn Glu
260 265 270

Ile Ala Thr Ala Leu Arg Lys Leu Gly Ser Asn Val Phe Val Leu Asp
275 280 285

Thr Asn Phe Ser Ala Asp Leu Thr Ile Ile Glu Glu Gly His Glu Leu
290 295 300

Ile Glu Arg Leu Tyr Arg Asn Val Thr Gly Lys Lys Leu Leu Gly Gly
305 310 315 320

Asp His Met Pro Ile Asp Leu Pro Met Leu Thr Ser Cys Cys Pro Gly
325 330 335

Trp Ile Met Phe Ile Glu Lys Asn Tyr Pro Asp Leu Leu Asn Asn Leu
340 345 350

Ser Thr Cys Lys Ser Pro Gln Gly Met Leu Gly Ala Leu Ile Lys Gly
355 360 365

Tyr Trp Ala Lys Asn Ile Lys Lys Met Asp Pro Lys Asp Ile Val Ser
370 375 380

Val Ser Ile Met Pro Cys Thr Ala Lys Lys Ala Glu Lys Glu Arg Pro
385 390 395 400

Gln Leu Arg Gly Asp Glu Gly Tyr Lys Asp Val Asp Tyr Ile Leu Thr
405 410 415

Thr Arg Glu Leu Ala Lys Met Leu Lys Gln Ser Asn Ile Asp Leu Ala
420 425 430

Lys Met Glu Pro Thr Pro Phe Asp Lys Val Met Ser Glu Gly Thr Gly

440 445

Ala Ala Val Ile Phe Gly Val Thr
450 455

<210> 45
<211> 369
<212> PRT
<213> Trichomonas vaginalis

<400> 45

Cys Asp Gly Lys Trp Leu Ala Pro Ala Cys Val Thr Thr Val Trp Asp
1 5 10 15

Gly Leu Lys Ile Asp Thr Lys Ser Lys Met Val Lys Glu Ser Val Glu
20 25 30

Asn Asn Leu Lys Glu Leu Leu Asp Cys His Asp Glu Thr Cys Ser Ser
35 40 45

Cys Val Ala Asn His Arg Cys Gln Phe Arg Asp Met Asn Val Ala Tyr
50 55 60

Ser Ile Lys Ala Glu Thr Lys Glu Glu Cys Ser Glu Glu Gly Ile Asp
65 70 75 80

Glu Ser Thr Asn Ser Ile Arg Leu Asp Thr Ser Lys Cys Val Leu Cys
85 90 95

Gly Arg Cys Ile Arg Ala Cys Glu Glu Val Ala Gly Gln Ser Ala Ile
100 105 110

Ile Phe Gly Asn Arg Ala Lys His Met Arg Ile Gln Pro Thr Phe Gly
115 120 125

Gln Thr Leu Gln Asp Thr Ser Cys Ile Lys Cys Gly Gln Cys Thr Leu
130 135 140

Tyr Cys Pro Val Gly Ala Ile Thr Glu Lys Ser Gln Val Lys Gln Ala
145 150 155 160

Leu Asp Ile Leu Ser Asn Lys Gly Lys Lys Ile Ser Val Ile Gln Val
165 170 175

Ala Pro Ala Val Arg Val Ala Leu Ser Glu Ala Phe Gly Tyr Lys Glu
180 185 190

Gly Ser Val Thr Thr Gly Lys Met Val Ser Ala Leu Lys Ala Leu Gly
195 200 205

Phe Asp Tyr Val Tyr Asp Thr Asn Tyr Ser Ala Asp Leu Thr Ile Val
210 215 220

Glu Glu Ala Gly Glu Leu Val Gln Arg Leu Lys Asn Pro Asn Ala Val
225 230 235 240

Phe Pro Met Phe Thr Ser Cys Cys Pro Ala Trp Val Asn Tyr Val Glu

245 250 255

Gln Ser Ala Pro Asp Phe Ile Pro Asn Leu Ser Ser Cys Arg Ser Pro
260 265 270

Gln Gly Met Leu Ser Ser Leu Val Lys Asn Tyr Leu Pro Lys Val Leu
275 280 285

Asn Ile Pro Val Glu Asp Val Leu Asn Phe Ser Ile Met Pro Cys Thr
290 295 300

Ala Lys Lys Asp Glu Ile Glu Arg Pro Glu Leu Arg Thr Lys Asp Gly
305 310 315 320

His Lys Glu Thr Asp Met Val Leu Thr Val Arg Glu Leu Val Glu Met
325 330 335

Ile Lys Leu Ser Gly Ile Asp Phe Asn Asn Leu Pro Asp Thr Pro Phe
340 345 350

Asp Ser Ile Phe Gly Phe Gly Ser Gly Ala Gly Gln Ile Phe Ala Ala
355 360 365

Thr

<210> 46
<211> 464
<212> PRT
<213> Entamoeba histolytica

<400> 46

Arg Leu His Thr Val Thr Gly His Asp His Asn His Ser Ile Gln Phe
1 5 10 15

Asp Trp Ser Lys Cys Met Gly Cys Gly Met Cys Ala Thr Lys Cys Thr
20 25 30

Phe Gly Val Leu Val Lys Gln Pro Pro Lys Ile Pro Pro Phe Val Gln
35 40 45

Pro Asn Arg Glu Lys Leu Ser Gln Glu Asn Thr Asp Lys Thr Arg Val
50 55 60

Leu Ile Asp Glu Ser Glu Cys Thr Gly Cys Gly Gln Cys Ser Leu Val
65 70 75 80

Cys Asn Phe Gly Ser Ile Thr Pro Ile Asp His Leu Val Asp Thr Phe
85 90 95

Lys Ala Lys Glu Ala Gly Lys Lys Leu Val Ala Met Ile Ala Pro Ser
100 105 110

Thr Arg Leu Gly Val Ala Glu Ala Met Gly Met Pro Ile Gly Ser Thr
115 120 125

Ala Met Ala Gln Leu Val His Cys Leu Arg Leu Ile Gly Phe Asp Tyr

130 135 140

Val Phe Asp Val Asp Ala Gly Ala Asp Lys Thr Thr Met Asp Asp Tyr
145 150 155 160

Ala Glu Val Ile Glu Met Lys Lys Glu Gly Lys Gly Pro Ala Ile Thr
165 170 175

Ser Cys Cys Pro Ala Trp Ile Glu Leu Val Glu Lys Glu Tyr Pro Asp
180 185 190

Leu Ile Pro Asn Val Ser Thr Ala Arg Ser Pro Ile Gly Cys Leu Ala
195 200 205

Gly Cys Ile Lys Arg Gly Trp Ala Lys Asp Val Gly Ile Ala Val Glu
210 215 220

Asp Leu Tyr Thr Val Gly Ile Met Pro Cys Ile Ala Lys Lys Thr Glu
225 230 235 240

Ser Gln Arg Gln Gln Ile His Gln Asp Tyr Asp Ala Ser Cys Thr Ser
245 250 255

Asn Glu Ile Ala Ala Tyr Phe Lys Lys His Leu Pro Pro Glu Glu Cys
260 265 270

Lys Phe Thr Gln Glu Arg Glu Glu Ala Leu Ala Lys Thr Glu Asp Gly
275 280 285

Gln Cys Asp Leu Pro Phe Arg Arg Ile Ser Gly Gly Ser Asn Ile Phe
290 295 300

Gly Arg Thr Gly Gly Val Cys Glu Thr Val Leu Arg Val Ile Ala Arg
305 310 315 320

Asn Ala Gly Val Asp Trp Asn Ser Cys Thr Val Asn Lys Glu Glu Thr
325 330 335

Phe Lys His Ala Ala Ser Gly Ser Thr Met Thr Asn Leu Ser Val Asp
340 345 350

Ile Gly Gly Thr Ile Ile Thr Gly Ala Val Cys His Gly Gly Tyr Ala
355 360 365

Ile Arg His Ala Cys Glu Leu Ile Arg Lys Gly Glu Leu Lys Val Asp
370 375 380

Val Val Glu Met Met Ala Cys Val Gly Gly Cys Leu Gly Gly Ala Gly
385 390 395 400

Gln Pro Lys Ile Pro Pro Ala Lys Lys Leu Glu Met Asp Lys Arg Arg
405 410 415

Val Met Leu Asp Ile Leu Asp Gln Gln Thr Asp Ile Arg Ala Ala Asn
420 425 430

Glu Asn Thr Asp Val Leu Gly Trp Ile Asp Lys His Phe Asp His Gln
435 440 445

Gly Ala His Gln His Leu His Thr Tyr Phe Thr Pro Arg Tyr Gln Asn
450 455 460

<210> 47
<211> 474
<212> PRT
<213> Giardia intestinalis

<400> 47

Met Pro Pro Lys Pro Gln His Asp Val Thr Gly Val Asp Ser Asn Asn
1 5 10 15

Ala Ile Met Ile Asp Tyr Ala Lys Cys Ile Gly Cys Asn Met Cys Ile
20 25 30

Lys Ala Cys Asp Val Gln Gly Ile Gly Val Tyr Lys Gln Asn Glu Lys
35 40 45

Pro Lys Tyr Pro Pro Ile Val Lys Leu Ser Thr Leu Phe Asn Ser Asp
50 55 60

Cys Ile Gly Cys Gly Gln Cys Ala Thr Ile Cys Pro Val Asp Ala Ile
65 70 75 80

Ala Pro Lys Asn Asn Leu Glu Ile Tyr Lys Gly Glu Ser Ala Ser Lys
85 90 95

Lys Val Arg Val Ala Leu Ile Ala Pro Ser Thr Arg Val Ala Phe Gly
100 105 110

Asp Val Phe Gly Leu Pro Ile Gly Thr Asn Thr Ile Tyr Ser Leu Ile
115 120 125

Arg Met Leu Lys Gln Tyr Leu Gly Phe Asp Tyr Val Phe Asp Val Asn
130 135 140

Phe Gly Ala Asp Glu Thr Thr Val Ile Asp Thr Gln Glu Leu Leu His
145 150 155 160

Phe Lys His Glu Gly Arg Gly Pro Val Phe Thr Ser Cys Cys Pro Ala
165 170 175

Trp Val Asn Leu Cys Glu Met Lys Tyr Pro Glu Leu Leu Pro Gln Val
180 185 190

Ser Thr Ala Lys Ser Cys Val Ala Met Val Ala Thr Leu Val Lys Arg
195 200 205

Arg Trp Val Gln Glu His Leu Ile Pro Lys Gly Ile Val Asp Ser Val
210 215 220

Asp Asp Val Tyr Val Ala Asp Ile Met Pro Cys Thr Ala Lys Lys Asp
225 230 235 240

Glu Ser Met Arg Pro Gln Leu Asn Arg Asp Val Asp Ile Cys Leu Thr
245 250 255

Val Arg Glu Val Ala Glu His Leu Tyr Phe Leu His Gly Ala Arg Leu
260 265 270

Thr Leu Glu Glu Val Glu Ala Asp Ala Leu Val Leu Arg Pro Gly Arg
275 280 285

Ser Thr Gln Lys Lys Trp Asp Phe Asp Ala Pro Phe Asn Thr Val Ser
290 295 300

Gly Gly Ser His Ile Phe Gly Lys Thr Gly Gly Val Ala Glu Thr Cys
305 310 315 320

Leu Arg Phe Ile Ser Tyr Met Lys Lys Ser Pro Ile Glu Asn Val Lys
325 330 335

Glu Glu Leu Leu Lys Glu Phe Lys Thr Pro Gly Gln Leu Val Gln Thr
340 345 350

Val Lys Leu Val Ser Cys Glu Ile Ala Gly Glu Thr Tyr Arg Ala Leu
355 360 365

Ile Ala His Gly Gly Ser Ala Ile Asn Ala Ala Ala Arg Met Val Leu
370 375 380

Asn Lys Glu Val Glu Cys Asp Val Val Glu Gln Met Ala Cys Pro Gly
385 390 395 400

Gly Cys Gln Asn Gly Gly Gly Met Pro Lys Ile Lys Gly Lys Lys Glu
405 410 415

Ala Val Leu Thr Arg Ala Ser Thr Leu Asp Ile Leu Asp Gly Lys Glu
420 425 430

Arg Phe Ala Ser Ala Gly Glu Asn Lys Thr Leu Trp Gly Phe Asn Gly
435 440 445

Cys Leu Thr Glu His Glu Ala His Glu Leu Leu His Thr His Tyr Gln
450 455 460

His Arg Pro Val Glu Ser Leu Leu Pro Gln
465 470

<210> 48
<211> 844
<212> PRT
<213> Desulfitobacterium hafniense

<400> 48

Met Val Lys Ile Ile Ser Ile Thr Asn Asn Ala Lys Arg Gln Gly Lys
1 5 10 15

Gly Thr Ser Arg Lys Glu Lys Gln Ala Met Lys Glu Val Thr Lys Gln
20 25 30

Gln Arg Ile Arg Val Thr Val Asn Gly Arg Gln Met Glu Val Tyr Gly
35 40 45

Asp Leu Thr Ile Leu Gln Ala Leu Leu Gln Glu Asp Ile His Ile Pro
50 55 60

His Leu Cys Tyr Asp Ile Arg Leu Glu Arg Ser Asn Gly Asn Cys Gly
65 70 75 80

Leu Cys Val Val Glu Leu Gly Glu Gly Ser Glu Gln Gln Asp Val Lys
85 90 95

Ala Cys His Thr Pro Ile Gln Glu Gly Met Ile Ile His Thr Asn Ser
100 105 110

Pro Arg Leu Glu His Tyr Arg Lys Ile Arg Leu Glu Gln Ile Leu Ala
115 120 125

Asp His Asn Ala Asp Cys Val Ala Pro Cys Val Met Thr Cys Pro Ala
130 135 140

Asn Ile Asp Ile Gln Ser Tyr Leu Ser His Ala Gly Asn Gly Asn Phe
145 150 155 160

Glu Thr Ala Ile Lys Val Ile Lys Glu Arg Asn Pro Phe Pro Ile Val
165 170 175

Cys Gly Arg Val Cys Pro His Ser Cys Glu Ala Gln Cys Arg Arg Asn
180 185 190

Leu Ile Asp Glu Pro Val Ala Ile Asn His Val Lys Arg Phe Ile Ala
195 200 205

Asp Trp Asp Ile Ala His Glu Gln Pro Trp Ala Pro Arg Lys Lys Ala
210 215 220

Ala Thr Gly Lys Lys Ile Ala Val Val Gly Ala Gly Ser Ser Gly Leu
225 230 235 240

Ser Ala Ala Tyr Tyr Ser Ala Ile Gln Gly His Asp Val Thr Val Phe
245 250 255

Glu Arg His Pro Arg Ala Gly Gly Met Met Arg Tyr Gly Ile Pro Glu
260 265 270

Tyr Arg Leu Pro Lys Glu Thr Leu Asp Arg Glu Ile Gly Leu Ile Ala
275 280 285

Asp Leu Gly Val Lys Ile Met Thr Asn Lys Ala Leu Gly Thr His Ile
290 295 300

Arg Leu Glu Asp Leu His Gln Asp Phe Asp Ala Val Tyr Leu Ala Ile
305 310 315 320

Gly Ser Trp Arg Ala Thr Pro Leu Gln Ile Glu Gly Asp Asn Leu Glu
325 330 335

Gly Val Trp Leu Gly Ile Asn Phe Leu Glu Gln Val Thr Lys Gly Ala
340 345 350

Asp Ile Lys Leu Gly Glu His Val Val Val Ile Gly Gly Gly Asn Thr
355 360 365

Ala Ile Asp Cys Ala Arg Thr Ala Leu Arg Lys Gly Ala Gly Ser Val
370 375 380

Lys Leu Val Tyr Arg Arg Thr Arg Glu Glu Met Pro Ala Glu Ser Tyr
385 390 395 400

Glu Val Glu Glu Ala Ile His Glu Gly Val Glu Met Tyr Phe Leu Thr
405 410 415

Ala Pro His Lys Ile Val Ala Glu Gly Gly Arg Lys Leu Leu His Cys
420 425 430

Ile Lys Met Thr Leu Gly Glu Pro Asp Arg Ser Gly Arg Arg Arg Pro
435 440 445

Ile Pro Ile Glu Gly Ser Glu Thr Ala Phe Glu Ala Asp Thr Ile Ile
450 455 460

Gly Ala Ile Gly Gln Ser Thr Asn Thr Gln Phe Leu Tyr His Asp Leu
465 470 475 480

Pro Val Lys Leu Asn Lys Trp Gly Asp Ile Glu Ile Asn Gly Lys Thr
485 490 495

Met Gln Thr Ser Glu Met Asn Ile Phe Ala Gly Gly Asp Cys Val Thr
500 505 510

Gly Pro Ala Thr Val Ile Gln Ala Val Ala Ala Gly Arg His Ala Ala
515 520 525

Glu Ala Met Asp Ser Phe Leu Met Lys Gly Tyr Val Lys Glu Gln Pro
530 535 540

Met Asp Tyr Ser Cys Ser Arg Gly Ser Leu Glu Asp Leu Pro Gln Trp
545 550 555 560

Glu Phe Glu Lys Ile Pro Arg Leu Lys Arg Ala Pro Met Pro Ala Leu
565 570 575

Pro Pro Ala Glu Arg Arg Asp Asn Phe Arg Glu Val Glu Thr Gly Leu
580 585 590

Ser Glu Glu Thr Ala Arg Ala Glu Ala Arg Arg Cys Leu Lys Cys Gly
595 600 605

Cys Tyr Glu Arg Tyr Asp Cys Asp Leu Arg Gln Glu Ala Ser Leu His
610 615 620

His Val Glu Phe Lys Lys Pro Val His Glu Arg Pro Tyr Ile Pro Ile
625 630 635 640

Val Glu Asp His Ser Ile Ile Ile Arg Asp His Asn Lys Cys Ile Ser
645 650 655

Cys Gly Arg Cys Ile Ala Ala Cys Ala Glu Val Glu Gly Pro Asp Ile
660 665 670

Leu Ser Phe Tyr Met Lys His Gly Arg Gln Leu Val Gly Thr Lys Ser
675 680 685

Gly Leu Pro Leu Asp Gln Thr Asp Cys Val Ser Cys Gly Gln Cys Val
690 695 700

Asn Ala Cys Pro Cys Gly Ala Leu Asp Tyr Arg Ser Glu Ile Gly Arg
705 710 715 720

Val Phe Arg Ala Ile Asn Asp Pro Gly Lys Thr Thr Val Ala Phe Val
725 730 735

Ala Pro Ala Val Arg Ser Val Val Ser Ser Gln Tyr Gly Val Ser Tyr
740 745 750

Gln Glu Ala Ser Arg Phe Ile Ala Gly Leu Leu Lys Lys Ile Gly Phe
755 760 765

Asp Lys Val Phe Asp Phe Thr Phe Ala Ala Asp Leu Thr Ile Val Glu
770 775 780

Glu Thr Thr Glu Phe Leu Thr Arg Leu Gln Ser His Lys Pro Ile Pro
785 790 795 800

Gln Phe Thr Ser Cys Cys Pro Gly Trp Val Asn Phe Val Glu Arg Arg
805 810 815

Tyr Pro Glu Ile Ile Pro Tyr Leu Ser Ser Cys Lys Ser Pro Gln Met
820 825 830

Met Met Gly Ala Thr Val Lys Ile Thr Leu Arg Asn
835 840

<210> 49
<211> 119
<212> PRT
<213> Nyctotherus velox

<400> 49

Ile Leu Phe Met Glu Lys Asn Tyr Pro Asp Met Leu Asn His Leu Ser
1 5 10 15

Thr Cys Lys Ser Pro Gln Gly Met Leu Gly Ala Leu Ile Lys Gly Tyr
20 25 30

Trp Ala Lys Asn Val Lys Lys Ile Asp Pro Lys Asp Val Val Ser Val
35 40 45

Ser Ile Met Pro Cys Thr Ala Lys Lys Glu Glu Lys Asp Arg Ile Thr
50 55 60

Leu Lys Ser Asp Glu Gly Tyr Asn Asn Val Asp Tyr Val Leu Thr Thr
65 70 75 80

Arg Glu Leu Ala Lys Met Phe Lys Gln Ser Asn Ile Asp Pro Ser Lys
85 90 95

Leu Pro Pro Thr Gln Phe Asp Asn Val Met Ser Glu Gly Thr Gly Ala
100 105 110

Ala Val Ile Phe Gly Val Thr
115

<210> 50
<211> 476
<212> PRT
<213> Oryza sativa

<400> 50

Met Ala Ser Ser Ser Ser Ser Ala Ser Ser Arg Phe Ser Pro Ala Leu
1 5 10 15

Gln Ala Ser Asp Leu Asn Asp Phe Ile Ala Pro Ser Gln Asp Cys Ile
20 25 30

Ile Ser Leu Asn Lys Gly Pro Ser Ala Arg Arg Leu Pro Ile Lys Gln
35 40 45

Lys Glu Ile Ala Val Ser Thr Asn Pro Pro Glu Glu Ala Val Lys Ile
50 55 60

Ser Leu Lys Asp Cys Leu Ala Cys Ser Gly Cys Ile Thr Ser Ala Glu
65 70 75 80

Thr Val Met Leu Glu Lys Gln Ser Leu Gly Asp Phe Ile Thr Arg Ile
85 90 95

Asn Ser Asp Lys Ala Val Ile Val Ser Val Ser Pro Gln Ser Arg Ala
100 105 110

Ser Leu Ala Ala Phe Phe Gly Leu Ser Gln Ser Gln Val Phe Arg Lys
115 120 125

Leu Thr Ala Leu Phe Lys Ser Met Gly Val Lys Ala Val Tyr Asp Thr
130 135 140

Ser Ser Ser Arg Asp Leu Ser Leu Ile Glu Ala Cys Ser Glu Phe Val
145 150 155 160

Thr Arg Tyr His Gln Asn Gln Leu Ser Ser Gly Lys Glu Ala Gly Lys
165 170 175

Asn Leu Pro Met Leu Ser Ser Ala Cys Pro Gly Trp Ile Cys Tyr Ala
180 185 190

Glu Lys Thr Leu Gly Ser Phe Ile Leu Pro Tyr Ile Ser Ala Val Lys
195 200 205

```
Ser Pro Gln Gln Ala Ile Gly Ala Ala Ile Lys His His Met Val Gly
    210                 215                 220

Lys Leu Gly Leu Lys Pro His Asp Val Tyr His Val Thr Val Met Pro
225                 230                 235                 240

Cys Tyr Asp Lys Lys Leu Glu Ala Val Arg Asp Asp Phe Val Phe Ser
                245                 250                 255

Val Glu Asp Lys Asp Val Thr Glu Val Asp Ser Val Leu Thr Thr Gly
            260                 265                 270

Glu Val Leu Asp Leu Ile Gln Ser Arg Ser Val Asp Phe Lys Thr Leu
        275                 280                 285

Glu Glu Ser Pro Met Asp Arg Leu Leu Thr Asn Val Asp Asp Asp Gly
    290                 295                 300

Gln Leu Tyr Gly Val Ser Gly Gly Ser Gly Gly Tyr Ala Glu Thr Val
305                 310                 315                 320

Phe Arg His Ala Ala His Val Leu Phe Asp Arg Lys Ile Glu Gly Ser
                325                 330                 335

Val Asp Phe Arg Ile Leu Arg Asn Ser Asp Phe Arg Glu Val Thr Leu
            340                 345                 350

Glu Val Glu Gly Lys Pro Val Leu Lys Phe Ala Leu Cys Tyr Gly Phe
        355                 360                 365

Arg Asn Leu Gln Asn Ile Ile Arg Lys Ile Lys Met Gly Lys Cys Glu
    370                 375                 380

Tyr His Phe Ile Glu Val Met Ala Cys Pro Ser Gly Cys Leu Asn Gly
385                 390                 395                 400

Gly Gly Gln Ile Lys Pro Ala Lys Gly Gln Ser Ala Lys Asp Leu Ile
                405                 410                 415

Gln Leu Leu Glu Asp Val Tyr Ile Gln Asp Val Ser Val Ser Asn Pro
            420                 425                 430

Phe Glu Asn Pro Ile Ala Lys Arg Leu Tyr Asp Glu Trp Leu Gly Gln
        435                 440                 445

Pro Gly Ser Glu Asn Ala Lys Lys Tyr Leu His Thr Lys Tyr His Pro
    450                 455                 460

Val Val Lys Ser Val Ala Ser Gln Leu Gln Asn Trp
465                 470                 475
```

<210> 51
<211> 114
<212> PRT
<213> Psalteriomonas lanterna

<400> 51

Ile Asn Leu Val Glu Lys His Tyr Pro Glu Tyr Leu Pro Asn Leu Ser
1 5 10 15

Ser Cys Arg Ser Pro Gln Gly Met Leu Ser Ser Leu Ile Lys Asn Tyr
20 25 30

Trp Ala Lys Lys Met Gly Ile Glu Pro Lys Asp Val Val Val Val Ser
35 40 45

Phe Met Pro Cys Gly Ala Lys Lys Asp Glu Ile Lys Arg Pro Gln Leu
50 55 60

Lys Gly Glu Thr Asp Tyr Val Leu Thr Thr Arg Glu Leu Gly Lys Leu
65 70 75 80

Phe Lys Met Gly Gly Leu Asn Asp Leu Ser Val Leu Glu Pro Val Lys
85 90 95

Tyr Asp Asp Pro Leu Gly Glu Ser Thr Gly Ala Ala Val Ile Phe Gly
100 105 110

Ala Thr

<210> 52
<211> 119
<212> PRT
<213> Nyctotherus ovalis

<400> 52

Ile Met Phe Met Glu Lys Asn Tyr Pro Asp Met Leu Asn His Leu Ser
1 5 10 15

Thr Cys Lys Ser Pro Gln Gly Met Leu Gly Ala Leu Ile Lys Gly Tyr
20 25 30

Trp Ala Lys Asn Ile Lys Lys Met Asp Pro Lys Asp Ile Val Ser Val
35 40 45

Ser Ile Met Pro Cys Thr Ala Lys Lys Ala Glu Lys Glu Arg Pro Gln
50 55 60

Leu Arg Gly Asp Glu Gly Tyr Lys Asp Val Asp Tyr Ile Leu Thr Thr
65 70 75 80

Arg Glu Leu Ala Lys Met Leu Lys Gln Ser Asn Ile Asp Leu Gly Lys
85 90 95

Met Glu Pro Thr Pro Phe Asp Lys Val Met Ser Glu Gly Thr Gly Ala
100 105 110

Ala Val Ile Phe Gly Val Thr
115

<210> 53
<211> 119

<212> PRT
<213> Nyctotherus ovalis

<400> 53

Ile Met Phe Met Glu Lys Asn Tyr Pro Asp Met Leu Asn His Leu Ser
1 5 10 15

Thr Cys Lys Ser Pro Gln Gly Met Leu Gly Ala Leu Ile Lys Gly Tyr
20 25 30

Trp Ala Lys Asn Val Lys Lys Met Asp Pro Lys Asp Ile Val Ser Val
35 40 45

Ser Ile Met Pro Cys Thr Ala Lys Lys Ala Glu Lys Glu Arg Pro Gln
50 55 60

Leu Arg Gly Asp Glu Gly Tyr Lys Asp Val Asp Tyr Ile Leu Thr Thr
65 70 75 80

Arg Glu Leu Ala Lys Met Leu Lys Gln Ser Asn Ile Asp Leu Gly Lys
85 90 95

Met Glu Pro Arg Pro Phe Asp Lys Val Met Ser Glu Gly Thr Gly Ala
100 105 110

Ala Val Ile Phe Gly Val Thr
115

<210> 54
<211> 520
<212> PRT
<213> Rhodospirillum rubrum

<400> 54

Met Arg Pro Val Gln Arg Pro Arg Arg Trp Pro Gly Leu Arg Gln Arg
1 5 10 15

Leu Ser Pro Glu Arg Pro Val Asp Arg Arg Ser Arg Arg Arg Ser Gly
20 25 30

Ala Ala Arg Pro Gly Arg Arg Arg Gly Ser Gly Val Gln His Glu Ile
35 40 45

Leu Arg Ser Val Ser Gln Arg Asp Met Ser Met Ser Ile Gln Pro Thr
50 55 60

Val Thr Ile Asp Pro Glu Leu Cys Thr Gly Cys Gly Arg Cys Val Glu
65 70 75 80

Thr Cys Pro Val Gln Ala Ile Ala Gly Ser Arg Gly Lys Ala His Glu
85 90 95

Ile Glu Ala Ala Ala Cys Val Ser Cys Gly Arg Cys Val Ala Thr Cys
100 105 110

Ala Ala Phe Asp Ser Ile Phe Asp Ala Phe Pro Thr Pro Arg Pro Val
115 120 125

```
Arg Leu Lys Arg Arg Gly Leu Pro Gly Ser Leu Lys Glu Pro Leu Phe
    130                 135                 140

Ala Ala His Asp Pro Ser Arg Ile Glu Ala Val Arg Lys Ala Phe Ala
145                 150                 155                 160

Thr Pro Lys Arg Met Thr Val Met Gln Val Asp Thr Met Ala Cys Val
                165                 170                 175

Ala Leu Ala Glu Asp Phe Gly Leu Pro Pro Gly Ser Leu Ser Pro Leu
            180                 185                 190

Lys Ile Ala Ser Ala Ala Arg Gln Leu Gly Phe Asp Arg Val Tyr Arg
        195                 200                 205

Thr Ser Phe Pro Ala Gly Leu Ala Val Leu Glu Thr Ala His Glu Met
    210                 215                 220

Ala Ala Arg Leu Ala Asn Gly Gly Asn Leu Pro Val Ile Asn Ser Ser
225                 230                 235                 240

Cys Pro Ala Val Val Ala Phe Leu Glu Arg Arg Tyr Pro Glu Leu Leu
                245                 250                 255

His Tyr Leu Ser Thr Val Lys Ser Pro His Gln Ile Ala Gly Ala Leu
            260                 265                 270

Tyr Asn Ser Tyr Leu Ala Asp Ala Ala Asn Leu Ala Pro Ala Asn Ile
        275                 280                 285

His Lys Val Ser Val Val Ala Cys Leu Ser His Lys Ala Glu Ala Glu
    290                 295                 300

Arg Pro Glu Met Met Thr Cys Gly Cys Pro Asp Ile Asp Thr Val Leu
305                 310                 315                 320

Thr Ala Arg Glu Leu Ala Ile Leu Ile Lys Asp Ala Gly Ile Asp Val
                325                 330                 335

Pro Leu Leu Gly Asp Gly Glu Phe Asp Asn Asp Phe Pro Glu Ile Glu
            340                 345                 350

Gly Leu Asp Thr Leu Tyr Cys Ala Pro Gly Asp Val Ser Arg Ala Val
        355                 360                 365

Leu Gly Ala Gly Arg Trp Phe Leu Gly Gln Gly Glu Gly Val Gly Ala
    370                 375                 380

Pro Ala Gly Glu Thr Val Glu Val Leu Asp Glu Ala Thr Arg Leu Thr
385                 390                 395                 400

Arg Leu Ala Tyr Pro Gly Gly Thr Leu Gln Ala Leu Thr Val Ala Gly
                405                 410                 415

Phe Asp Lys Ala Val Pro Tyr Leu Glu Ala Ile Lys Ala Gly Arg Asn
            420                 425                 430
```

```
Ala Phe Gln Phe Leu Glu Ile Ala Ser Cys Pro Gln Gly Cys Ala Ser
        435                 440                 445

Gly Ala Gly Leu Pro Lys Val Leu Leu Glu Thr Glu Lys Pro Ala Arg
    450                 455                 460

Tyr Arg Ala Arg Ile Glu Asn Leu Pro Pro Ala Ala Pro Glu Ala Trp
465                 470                 475                 480

Ser Arg Leu Pro Gly His Pro Ser Ile Val Ala Leu Tyr Gly Gly Tyr
                485                 490                 495

Phe Gly Lys Ala Ile Gly Asp Lys Ser Asn Arg Arg Leu His Thr Gln
            500                 505                 510

Tyr Ala Glu Pro Ala Ala Ala Pro
        515                 520
```

<210> 55
<211> 240
<212> PRT
<213> Desulfitobacterium hafniense

<400> 55

```
Met Ala Val Glu Lys Leu Thr Gly Glu Val Leu Thr Asp Gln Leu Asp
1               5                   10                  15

Tyr Gln Glu Val Arg Gly Leu Gln Gly Ile Lys Glu Ala Ala Val Glu
            20                  25                  30

Ala Lys Gly Lys Lys Val Asn Val Ala Val Ile Ser Gly Leu His Asn
        35                  40                  45

Val Glu Pro Ile Leu Glu Lys Ile Ile Glu Gly Met Glu Val Gly Tyr
    50                  55                  60

Asp Leu Ile Glu Val Met Ala Cys Pro Gly Gly Cys Ile Cys Gly Ala
65                  70                  75                  80

Gly His Pro Val Pro Glu Lys Ile Asp Thr Leu Glu Lys Arg Gln Gln
                85                  90                  95

Val Leu Val Asn Ile Asp Gln Thr Ser Arg Tyr Arg Lys Ser Gln Glu
            100                 105                 110

Asn Pro Asp Ile Leu Arg Leu Tyr Asp Glu Tyr Tyr Gly Glu Ala Asn
        115                 120                 125

Ser Pro Leu Ala His Lys Leu Leu His Thr His Tyr Glu Ala Val Lys
    130                 135                 140

Arg Glu Pro Val Ala Lys His Asp Arg Arg Met Ala Asp Ser Ala Phe
145                 150                 155                 160

Val Thr His Glu Leu Thr Leu Cys Thr Cys Asp Lys Cys Thr Ala Gln
                165                 170                 175
```

Gly Ser Arg Glu Leu Phe Ala Ala Leu Ser Gly Lys Ile Arg Lys Leu
180 185 190

Lys Met Asp Ser Phe Val Thr Ala Arg Thr Ile Arg Leu Lys Glu Asn
195 200 205

His Pro Gly Gln Gly Val Tyr Ala Ala Ile Asp Gly Lys Leu Ile Glu
210 215 220

Thr Pro Val Glu Gln Leu Glu Gln Arg Ile Phe Gln His Leu Ile Arg
225 230 235 240

<210> 56
<211> 86
<212> PRT
<213> Desulfitobacterium hafniense

<400> 56

Met Val Ser Ile Val Pro Cys Ile Ala Lys Lys Tyr Glu Ala Ala Arg
1 5 10 15

Pro Glu Phe Arg Ser Glu Gly Ile Arg Asp Val Asp Ala Val Leu Thr
20 25 30

Ser Thr Glu Met Leu Glu Met Ala Asp Ile Lys Leu Ile Glu Pro Ala
35 40 45

Asp Val Glu Pro Gln Asp Phe Cys Glu Pro Tyr Lys Arg Val Ser Gly
50 55 60

Ala Gly Ile Leu Phe Gly Ala Ser Gly Gly Val Ala Lys Arg Pro Cys
65 70 75 80

Gly Trp Arg Trp Arg Asn
85

<210> 57
<211> 477
<212> PRT
<213> Drosophila melanogaster

<400> 57

Met Ser Arg Leu Ser Arg Ala Leu Gln Leu Thr Asp Ile Asp Asp Phe
1 5 10 15

Ile Thr Pro Ser Gln Ile Cys Ile Lys Pro Val Gln Ile Asp Lys Ala
20 25 30

Arg Ser Lys Thr Gly Ala Lys Ile Lys Ile Lys Gly Asp Gly Cys Phe
35 40 45

Glu Glu Ser Glu Ser Gly Asn Leu Lys Leu Asn Lys Val Asp Ile Ser
50 55 60

Leu Gln Asp Cys Leu Ala Cys Ser Gly Cys Ile Thr Ser Ala Glu Glu
65 70 75 80

Val Leu Ile Thr Gln Gln Ser Arg Glu Glu Leu Leu Lys Val Leu Gln
85 90 95

Glu Asn Ser Lys Asn Lys Ala Ser Glu Asp Trp Asp Asn Val Arg Thr
100 105 110

Ile Val Phe Thr Leu Ala Thr Gln Pro Ile Leu Ser Leu Ala Tyr Arg
115 120 125

Tyr Gln Ile Gly Val Glu Asp Ala Ala Arg His Leu Asn Gly Tyr Phe
130 135 140

Arg Ser Leu Gly Ala Asp Tyr Val Leu Ser Thr Lys Val Ala Asp Asp
145 150 155 160

Ile Ala Leu Leu Glu Cys Arg Gln Glu Phe Val Asp Arg Tyr Arg Glu
165 170 175

Asn Glu Asn Leu Thr Met Leu Ser Ser Ser Cys Pro Gly Trp Val Cys
180 185 190

Tyr Ala Glu Lys Thr His Gly Asn Phe Leu Leu Pro Tyr Val Ser Thr
195 200 205

Thr Arg Ser Pro Gln Gln Ile Met Gly Val Leu Val Lys Gln Ile Leu
210 215 220

Ala Asp Lys Met Asn Val Pro Ala Ser Arg Ile Tyr His Val Thr Val
225 230 235 240

Met Pro Cys Tyr Asp Lys Lys Leu Glu Ala Ser Arg Glu Asp Phe Phe
245 250 255

Ser Lys Ala Asn Asn Ser Arg Asp Val Asp Cys Val Ile Thr Ser Val
260 265 270

Glu Val Glu Gln Leu Leu Ser Glu Ala Gln Gln Pro Leu Ser Gln Tyr
275 280 285

Asp Leu Leu Asp Leu Asp Trp Pro Trp Ser Asn Val Arg Pro Glu Phe
290 295 300

Met Val Trp Ala His Glu Lys Thr Leu Ser Gly Gly Tyr Ala Glu His
305 310 315 320

Ile Phe Lys Tyr Ala Ala Lys His Ile Phe Asn Glu Asp Leu Lys Thr
325 330 335

Glu Leu Glu Phe Lys Gln Leu Lys Asn Arg Asp Phe Arg Glu Ile Ile
340 345 350

Leu Lys Gln Asn Gly Lys Thr Val Leu Lys Phe Ala Ile Ala Asn Gly
355 360 365

Phe Arg Asn Ile Gln Asn Leu Val Gln Lys Leu Lys Arg Glu Lys Val
370 375 380

Ser Asn Tyr His Phe Val Glu Val Met Ala Cys Pro Ser Gly Cys Ile
385 390 395 400

Asn Gly Gly Ala Gln Ile Arg Pro Thr Thr Gly Gln His Val Arg Glu
405 410 415

Leu Thr Arg Lys Leu Glu Glu Leu Tyr Gln Asn Leu Pro Arg Ser Glu
420 425 430

Pro Glu Asn Ser Leu Thr Lys His Ile Tyr Asn Asp Phe Leu Asp Gly
435 440 445

Phe Gln Ser Asp Lys Ser Tyr Asp Val Leu His Thr Arg Tyr His Asp
450 455 460

Val Val Ser Glu Leu Ser Ile Ser Leu Asn Ile Asn Trp
465 470 475

<210> 58
<211> 538
<212> PRT
<213> S. pombe

<400> 58

Met Ala Lys Leu Ser Val Asn Asp Leu Asn Asp Phe Leu Ser Pro Gly
1 5 10 15

Ala Val Cys Ile Lys Pro Ala Gln Val Lys Lys Gln Glu Ser Lys Asn
20 25 30

Asp Ile Arg Ile Asp Gly Asp Ala Tyr Tyr Glu Val Thr Lys Asp Thr
35 40 45

Gly Glu Thr Ser Glu Leu Gly Ile Ala Ser Ile Ser Leu Asn Asp Cys
50 55 60

Leu Ala Cys Ser Gly Cys Ile Thr Ser Ala Glu Thr Val Leu Val Asn
65 70 75 80

Leu Gln Ser Tyr Gln Glu Val Leu Lys His Leu Glu Ser Arg Lys Ser
85 90 95

Gln Glu Ile Leu Tyr Val Ser Leu Ser Pro Gln Val Arg Ala Asn Leu
100 105 110

Ala Ala Tyr Tyr Gly Leu Ser Leu Gln Glu Ile Gln Ala Val Leu Glu
115 120 125

Met Val Phe Ile Gly Lys Leu Gly Phe His Ala Ile Leu Asp Thr Asn
130 135 140

Ala Ser Arg Glu Ile Val Leu Gln Gln Cys Ala Gln Glu Phe Cys Asn
145 150 155 160

Ser Trp Leu Gln Ser Arg Ala His Lys Asn Gln Asn Gln Val Thr Asn
165 170 175

Ser Val Val Asn Glu His Pro Leu Ile Pro His Ser Thr Ser Gln Ile
180 185 190

Ser Gly Val His Ser Asn Thr Ser Ser Asn Ser Gly Ile Asn Glu Asn
195 200 205

Ala Val Leu Pro Ile Leu Ser Ser Ser Cys Pro Gly Trp Ile Cys Tyr
210 215 220

Val Glu Lys Thr His Ser Asn Leu Ile Pro Asn Leu Ser Arg Val Arg
225 230 235 240

Ser Pro Gln Gln Ala Cys Gly Arg Ile Leu Lys Asp Trp Ala Val Gln
245 250 255

Gln Phe Ser Met Gln Arg Asn Asp Val Trp His Leu Ser Leu Met Pro
260 265 270

Cys Phe Asp Lys Lys Leu Glu Ala Ser Arg Asp Glu Phe Ser Glu Asn
275 280 285

Gly Val Arg Asp Val Asp Ser Val Leu Thr Pro Lys Glu Leu Val Glu
290 295 300

Met Phe Lys Phe Leu Arg Ile Asp Pro Ile Glu Leu Thr Lys Asn Pro
305 310 315 320

Ile Pro Phe Gln Gln Ser Thr Asp Ala Ile Pro Phe Trp Tyr Pro Arg
325 330 335

Ile Thr Tyr Glu Glu Gln Ile Gly Ser Ser Ser Gly Gly Tyr Met Gly
340 345 350

Tyr Val Leu Ser Tyr Ala Ala Lys Met Leu Phe Gly Ile Asp Asp Val
355 360 365

Gly Pro Tyr Val Ser Met Asn Asn Lys Asn Gly Asp Leu Thr Glu Tyr
370 375 380

Thr Leu Arg His Pro Glu Thr Asn Glu Gln Leu Ile Ser Met Ala Thr
385 390 395 400

Cys Tyr Gly Phe Arg Asn Ile Gln Asn Leu Val Arg Arg Val His Gly
405 410 415

Asn Ser Ser Val Arg Lys Gly Arg Val Leu Leu Lys Lys Arg Val Arg
420 425 430

Ser Asn Ala Gln Asn Pro Thr Glu Glu Pro Ser Arg Tyr Asp Tyr Val
435 440 445

Glu Val Met Ala Cys Pro Gly Gly Cys Ile Asn Gly Gly Gly Gln Leu
450 455 460

Pro Phe Pro Ser Val Glu Arg Ile Val Ser Ala Arg Asp Trp Met Gln

465 470 475 480

Gln Val Glu Lys Leu Tyr Tyr Glu Pro Gly Thr Arg Ser Val Asp Gln
485 490 495

Ser Ala Val Ser Tyr Met Leu Glu Gln Trp Val Lys Asp Pro Thr Leu
500 505 510

Thr Pro Lys Phe Leu His Thr Ser Tyr Arg Ala Val Gln Thr Asp Asn
515 520 525

Asp Asn Pro Leu Leu Leu Ala Asn Lys Trp
530 535

<210> 59
<211> 119
<212> PRT
<213> Metopus contortus

<400> 59

Ile Ile Phe Ala Glu Lys Asn Tyr Pro Glu Met Val Asn His Leu Ser
1 5 10 15

Thr Thr Lys Ser Pro Met Gln Met Leu Ser Ser Leu Ser Lys Gly Tyr
20 25 30

Trp Ala Lys Glu Gly Lys Lys Ile Asp Pro Lys Asn Val Val Asn Val
35 40 45

Ala Ile Met Pro Cys Thr Ala Lys Lys Ala Trp Lys Glu Arg Pro Asp
50 55 60

Met Lys Ala Asp Asn Gly Asp Pro Val Thr Asp Tyr Val Leu Thr Thr
65 70 75 80

Arg Glu Leu Gly Thr Met Leu Arg Gln Ser Asn Ile Asn Pro Val Ser
85 90 95

Leu Pro Lys Thr Pro Phe Asp Lys Ile Met Gly Glu Ser Thr Gly Ala
100 105 110

Ala Val Ile Phe Gly Ala Thr
115

<210> 60
<211> 462
<212> PRT
<213> Mus musculus

<400> 60

Met Lys Cys Glu His Cys Thr Arg Lys Glu Cys Ser Lys Lys Ser Lys
1 5 10 15

Asn Asp Asp Gln Glu Asn Val Ser Ser Asp Gly Ala Gln Pro Ser Asp
20 25 30

Gly Ala Ser Pro Ala Lys Glu Ser Glu Glu Lys Gly Glu Phe His Lys
35 40 45

Leu Ala Asp Ala Lys Ile Phe Leu Ser Asp Cys Leu Ala Cys Asp Ser
50 55 60

Cys Val Thr Val Glu Glu Gly Val Gln Leu Ser Gln Gln Ser Ala Lys
65 70 75 80

Asp Phe Leu His Val Leu Asn Leu Asn Lys Arg Cys Asp Thr Ser Lys
85 90 95

His Arg Val Leu Val Val Ser Val Cys Pro Gln Ser Leu Pro Tyr Phe
100 105 110

Ala Ala Lys Phe Asn Leu Ser Val Thr Asp Ala Ser Arg Arg Leu Cys
115 120 125

Gly Phe Leu Lys Ser Leu Gly Val His Tyr Val Phe Asp Thr Thr Ile
130 135 140

Ala Ala Asp Phe Ser Ile Leu Glu Ser Gln Lys Glu Phe Val Arg Arg
145 150 155 160

Tyr His Gln His Ser Glu Glu Gln Arg Glu Leu Pro Met Leu Thr Ser
165 170 175

Ala Cys Pro Gly Trp Val Arg Tyr Ala Glu Arg Val Leu Gly Arg Pro
180 185 190

Ile Ile Pro Tyr Leu Cys Thr Ala Lys Ser Pro Gln Gln Val Met Gly
195 200 205

Ser Leu Val Lys Asp Tyr Phe Ala Arg Gln Gln Asn Leu Ser Pro Glu
210 215 220

Lys Ile Phe His Val Val Val Ala Pro Cys Tyr Asp Lys Lys Leu Glu
225 230 235 240

Ala Leu Arg Glu Gly Leu Ser Thr Thr Leu Asn Gly Ala Arg Gly Thr
245 250 255

Asp Cys Val Leu Thr Ser Gly Glu Ile Ala Gln Ile Met Glu Gln Ser
260 265 270

Asp Leu Ser Val Lys Asp Ile Ala Val Asp Thr Leu Phe Gly Asp Met
275 280 285

Lys Glu Val Ala Val Gln Arg His Asp Gly Val Ser Ser Asp Gly His
290 295 300

Leu Ala His Val Phe Arg His Ala Ala Lys Glu Leu Phe Gly Glu His
305 310 315 320

Val Glu Glu Ile Thr Tyr Arg Ala Leu Arg Asn Lys Asp Phe His Glu
325 330 335

Val Thr Leu Glu Lys Asn Gly Glu Val Leu Leu Arg Phe Ala Ala Ala

340 345 350

Tyr Gly Phe Arg Asn Ile Gln Asn Met Ile Gln Lys Leu Lys Lys Gly
355 360 365

Lys Leu Pro Tyr His Phe Val Glu Val Leu Ala Cys Pro Arg Gly Cys
370 375 380

Leu Asn Gly Arg Gly Gln Ala Gln Thr Glu Asp Gly His Thr Asp Arg
385 390 395 400

Ala Leu Leu Gln Gln Met Glu Gly Ile Tyr Ser Gly Ile Pro Val Arg
405 410 415

Pro Pro Glu Ser Ser Thr His Val Gln Glu Leu Tyr Gln Glu Trp Leu
420 425 430

Glu Gly Thr Glu Ser Pro Lys Val Gln Glu Val Leu His Thr Ser Tyr
435 440 445

Gln Ser Leu Glu Pro Cys Thr Asp Gly Leu Asp Ile Lys Trp
450 455 460

<210> 61
<211> 457
<212> PRT
<213> Caenorhabditis elegans

<400> 61

Met Glu Asp Ser Gly Phe Ser Gly Val Val Arg Leu Ser Asn Val Ser
1 5 10 15

Asp Phe Ile Ala Pro Asn Leu Asp Cys Ile Ile Pro Leu Glu Thr Arg
20 25 30

Thr Val Glu Lys Lys Lys Glu Glu Ser Gln Val Asn Ile Arg Thr Lys
35 40 45

Lys Pro Lys Asp Lys Glu Ser Ser Lys Thr Glu Glu Lys Lys Ser Val
50 55 60

Lys Ile Ser Leu Ala Asp Cys Leu Ala Cys Ser Gly Cys Ile Thr Ser
65 70 75 80

Ala Glu Thr Val Leu Val Glu Glu Gln Ser Phe Gly Arg Val Tyr Glu
85 90 95

Gly Ile Gln Asn Ser Lys Leu Ser Val Val Thr Val Ser Pro Gln Ala
100 105 110

Ile Thr Ser Ile Ala Val Lys Ile Gly Lys Ser Thr Asn Glu Val Ala
115 120 125

Lys Ile Ile Ala Ser Phe Phe Arg Arg Leu Gly Val Lys Tyr Val Ile
130 135 140

Asp Ser Ser Phe Ala Arg Lys Phe Ala His Ser Leu Ile Tyr Glu Glu

145 150 155 160

Leu Ser Thr Thr Pro Ser Thr Ser Arg Pro Leu Leu Ser Ser Ala Cys
165 170 175

Pro Gly Phe Val Cys Tyr Ala Glu Lys Ser His Gly Glu Leu Leu Ile
180 185 190

Pro Lys Ile Ser Lys Ile Arg Ser Pro Gln Ala Ile Ser Gly Ala Ile
195 200 205

Ile Lys Gly Phe Leu Ala Lys Arg Glu Gly Leu Ser Pro Cys Asp Val
210 215 220

Phe His Ala Ala Val Met Pro Cys Phe Asp Lys Lys Leu Glu Ala Ser
225 230 235 240

Arg Glu Gln Phe Lys Val Asp Gly Thr Asp Val Arg Glu Thr Asp Cys
245 250 255

Val Ile Ser Thr Ala Glu Leu Leu Glu Glu Ile Ile Lys Leu Glu Asn
260 265 270

Asp Glu Ala Gly Asp Val Glu Asn Arg Ser Glu Glu Glu Gln Trp Leu
275 280 285

Ser Ala Leu Ser Lys Gly Ser Val Ile Gly Asp Asp Gly Gly Ala Ser
290 295 300

Gly Gly Tyr Ala Asp Arg Ile Val Arg Asp Phe Val Leu Glu Asn Gly
305 310 315 320

Gly Gly Ile Val Lys Thr Ser Lys Leu Asn Lys Asn Met Phe Ser Thr
325 330 335

Thr Val Glu Ser Glu Ala Gly Glu Ile Leu Leu Arg Val Ala Lys Val
340 345 350

Tyr Gly Phe Arg Asn Val Gln Asn Leu Val Arg Lys Met Lys Thr Lys
355 360 365

Lys Glu Lys Thr Asp Tyr Val Glu Ile Met Ala Cys Pro Gly Gly Cys
370 375 380

Ala Asn Gly Gly Gly Gln Ile Arg Tyr Glu Thr Met Asp Glu Arg Glu
385 390 395 400

Glu Lys Leu Ile Lys Val Glu Ala Leu Tyr Glu Asp Leu Pro Arg Gln
405 410 415

Asp Asp Glu Glu Thr Trp Ile Lys Val Arg Glu Glu Trp Glu Lys Leu
420 425 430

Asp Lys Asn Tyr Arg Asn Leu Leu Phe Thr Asp Tyr Arg Pro Val Glu
435 440 445

Thr Asn Val Ala Gln Val Leu Lys Trp
450 455

<210> 62
<211> 462
<212> PRT
<213> Mus musculus

<400> 62

Met Lys Cys Glu His Cys Thr Arg Lys Glu Cys Ser Lys Lys Ser Lys
1 5 10 15

Thr Asp Asp Gln Glu Asn Val Ser Ser Asp Gly Ala Gln Pro Ser Asp
20 25 30

Gly Ala Ser Pro Ala Lys Glu Ser Glu Glu Lys Gly Glu Phe His Lys
35 40 45

Leu Ala Asp Ala Lys Ile Phe Leu Ser Asp Cys Leu Ala Cys Asp Ser
50 55 60

Cys Val Thr Val Glu Glu Gly Val Gln Leu Ser Gln Gln Ser Ala Lys
65 70 75 80

Asp Phe Leu His Val Leu Asn Leu Asn Lys Arg Cys Asp Thr Ser Lys
85 90 95

His Arg Val Leu Val Val Ser Val Cys Pro Gln Ser Leu Pro Tyr Phe
100 105 110

Ala Ala Lys Phe Asn Leu Ser Val Thr Asp Ala Ser Arg Arg Leu Cys
115 120 125

Gly Phe Leu Lys Ser Leu Gly Val His Tyr Val Phe Asp Thr Thr Ile
130 135 140

Ala Ala Asp Phe Ser Ile Leu Glu Ser Gln Lys Glu Phe Val Arg Arg
145 150 155 160

Tyr His Gln His Ser Glu Glu Gln Arg Glu Leu Pro Met Leu Thr Ser
165 170 175

Ala Cys Pro Gly Trp Val Arg Tyr Ala Glu Arg Val Leu Gly Arg Pro
180 185 190

Ile Ile Pro Tyr Leu Cys Thr Ala Lys Ser Pro Gln Gln Val Met Gly
195 200 205

Ser Leu Val Lys Asp Tyr Phe Ala Arg Gln Gln Asn Leu Ser Pro Glu
210 215 220

Lys Ile Phe His Val Val Val Ala Pro Cys Tyr Asp Lys Lys Leu Glu
225 230 235 240

Ala Leu Arg Glu Gly Leu Ser Thr Thr Leu Asn Gly Ala Arg Gly Thr
245 250 255

PCT/US2005/001983
Asp Cys Val Leu Thr Ser Gly Glu Ile Ala Gln Ile Met Glu Gln Ser
260 265 270

Asp Leu Ser Val Lys Asp Ile Ala Val Asp Thr Leu Phe Gly Asp Met
275 280 285

Lys Glu Val Ala Val Gln Arg His Asp Gly Val Ser Ser Asp Gly His
290 295 300

Leu Ala His Val Phe Arg His Ala Ala Lys Glu Leu Phe Gly Glu His
305 310 315 320

Val Glu Glu Ile Thr Tyr Arg Ala Leu Arg Asn Lys Asp Phe His Glu
325 330 335

Val Thr Leu Glu Lys Asn Gly Glu Val Leu Leu Arg Phe Ala Ala Ala
340 345 350

Tyr Gly Phe Arg Asn Ile Gln Asn Met Ile Gln Lys Leu Lys Lys Gly
355 360 365

Lys Leu Pro Tyr His Phe Val Glu Val Leu Ala Cys Pro Arg Gly Cys
370 375 380

Leu Asn Gly Arg Gly Gln Ala Gln Thr Glu Asp Gly His Thr Asp Arg
385 390 395 400

Ala Leu Leu Gln Gln Met Glu Gly Ile Tyr Ser Gly Ile Pro Val Arg
405 410 415

Pro Pro Glu Ser Ser Thr His Val Gln Glu Leu Tyr Gln Glu Trp Leu
420 425 430

Glu Gly Thr Glu Ser Pro Lys Val Gln Glu Val Leu His Thr Ser Tyr
435 440 445

Gln Ser Leu Glu Pro Cys Thr Asp Gly Leu Asp Ile Lys Trp
450 455 460

<210> 63
<211> 119
<212> PRT
<213> Neocallimastix

<400> 63

Ile Met Phe Ala Glu Lys Asn Phe Pro Asp Met Val Asn Asn Leu Ser
1 5 10 15

Thr Thr Lys Ser Pro Met Gln Met Leu Ser Ser Leu Thr Lys Gly Tyr
20 25 30

Trp Ala Lys Asp Ile Lys Lys Ile Asn Pro Lys Asp Val Val Asn Val
35 40 45

Ala Ile Met Pro Cys Thr Ala Lys Lys Gln Glu Lys Asp Arg Pro Gly
50 55 60

Met Lys Thr Asp Glu Gly Asp Lys Val Thr Asp Phe Val Leu Thr Thr
65 70 75 80

Arg Glu Leu Gly Met Met Leu Arg Gln Ala Asn Ile Asp Pro Thr Lys
85 90 95

Leu Pro Gly Thr Lys Phe Asp Lys Val Met Gly Glu Ser Thr Gly Ala
100 105 110

Ala Val Ile Phe Gly Ala Thr
115

<210> 64
<211> 119
<212> PRT
<213> Nyctotherus ovalis

<400> 64

Ile Ile Phe Met Glu Lys Asn Tyr Pro Asp Met Leu Ser His Leu Ser
1 5 10 15

Thr Cys Lys Ser Pro Gln Gly Met Leu Gly Ala Leu Ile Lys Gly Tyr
20 25 30

Trp Ala Lys Lys Val Lys Lys Val Asp Pro Lys Asp Val Val Ser Val
35 40 45

Ser Ile Met Pro Cys Thr Ala Lys Lys Ala Glu Lys Glu Arg Pro Gln
50 55 60

Leu Arg Gly Asp Glu Gly Phe Lys Asp Val Asp Tyr Val Leu Thr Thr
65 70 75 80

Arg Glu Leu Ala Lys Met Leu Lys Gln Ser Asn Ile Asp Leu Gly Lys
85 90 95

Val Glu Pro Thr Pro Phe Asp Ala Val Met Ser Glu Gly Thr Gly Ala
100 105 110

Ala Val Ile Phe Gly Val Thr
115

<210> 65
<211> 490
<212> PRT
<213> Clostridium perfringens

<400> 65

Met Ala Ile Lys Asp Ala Asn Lys Gln Tyr Ile Lys Phe Asp Thr Ala
1 5 10 15

Val Gln Val Leu Lys Tyr Glu Val Leu Lys Arg Ile Ala Glu Lys Glu
20 25 30

Phe Asp Gly Thr Leu Asp Lys Glu Lys Leu Asn Ile Ala Lys Glu Ile
35 40 45

Val Asp Asp Leu Lys Pro Asn Val Arg Cys Cys Ile Tyr Lys Glu Arg

50 55 60

Ala Ile Val Glu Glu Arg Met Lys Leu Ala Leu Gly Gly His Glu Asn
65 70 75 80

Arg Glu Asn Met Ile Glu Val Ile Asp Ile Ala Cys Asp Glu Cys Pro
85 90 95

Val Asn Arg Phe Ile Val Thr Asp Ala Cys Arg Gly Cys Leu Ala Lys
100 105 110

Lys Cys Arg Asp Ser Cys Asn Phe Gly Ala Ile Ser Phe Asp Asn Arg
115 120 125

Lys Cys Lys Ile Asp Tyr Glu Lys Cys Lys Glu Cys Gly Lys Cys Lys
130 135 140

Glu Val Cys Pro Tyr Asn Ala Ile Ala Glu Val Lys Arg Pro Cys Met
145 150 155 160

Arg Ala Cys Ile Pro Lys Ala Leu Ser Tyr Asp Val Asp Ser Lys Lys
165 170 175

Ala Val Ile Asp Asp Ser Lys Cys Ile Gln Cys Gly Ala Cys Val Val
180 185 190

Asp Cys Pro Phe Gly Ala Ile Met Asp Lys Ser Tyr Leu Val Asp Val
195 200 205

Ile Arg Leu Leu Lys Asp Glu Lys Lys Val Tyr Ala Ile Val Ala Pro
210 215 220

Ala Ile Ser Ser Gln Phe Asn His Ser Lys Ile Gly Lys Val Ile Thr
225 230 235 240

Ala Ile Lys Lys Leu Gly Phe Glu Asp Val Phe Glu Ala Ala Leu Gly
245 250 255

Ala Asp Leu Val Ala Val His Glu Cys Asn Glu Phe Lys Glu Lys Gly
260 265 270

Glu Leu Asp Phe Met Thr Thr Ser Cys Cys Pro Ala Phe Val Ser Tyr
275 280 285

Ile Glu Lys Asn Tyr Pro Glu Leu Lys Glu Cys Ile Ser Asn Thr Val
290 295 300

Ser Pro Met Val Ala Met Ala Arg Leu Ile Lys Ser Gln Asn Lys Asp
305 310 315 320

Val Lys Thr Val Phe Ile Gly Pro Cys Ile Ala Lys Lys Thr Glu Ala
325 330 335

Lys Arg Asn Glu Val Ser Gly Asp Val Asp Tyr Val Leu Thr Phe Glu
340 345 350

Glu Leu Leu Ala Leu Leu Asp Ser Arg Asn Ile Lys Ile Asp Glu Cys
355 360 365

Glu Glu Ser Asp Thr Lys His Gly Ser Phe Tyr Gly Arg Leu Phe Ala
370 375 380

Arg Ser Gly Gly Val Thr Glu Ser Val Lys His Leu Ile Asp Ser Glu
385 390 395 400

Gly Ile Lys Val Asp Phe Arg Pro Ile Leu Gly Asp Gly Ile Lys Asp
405 410 415

Cys Asp Ile Lys Leu Arg Leu Ala Lys Leu Lys Arg Ala Gln Gly Asn
420 425 430

Phe Leu Glu Gly Met Ala Cys Lys Gly Gly Cys Ile Asn Gly Pro Gly
435 440 445

Ser Leu Asn His Asp Ile Lys Asn Ser Lys Glu Val Asp Lys Tyr Gly
450 455 460

Glu Leu Ser Ser Ser Glu Lys Ile Lys Asp Thr Leu Ala Asp Ile Lys
465 470 475 480

Phe Glu Asp Leu Asn Leu Ser Lys Asn Glu
485 490

<210> 66
<211> 456
<212> PRT
<213> Homo sapiens

<400> 66

Met Lys Cys Glu His Cys Thr Arg Lys Glu Cys Ser Lys Lys Thr Lys
1 5 10 15

Thr Asp Asp Gln Glu Asn Val Ser Ala Asp Ala Pro Ser Pro Ala Gln
20 25 30

Glu Asn Gly Glu Lys Gly Glu Phe His Lys Leu Ala Asp Ala Lys Ile
35 40 45

Phe Leu Ser Asp Cys Leu Ala Cys Asp Ser Cys Met Thr Ala Glu Glu
50 55 60

Gly Val Gln Leu Ser Gln Gln Asn Ala Lys Asp Phe Phe Arg Val Leu
65 70 75 80

Asn Leu Asn Lys Lys Cys Asp Thr Ser Lys His Lys Val Leu Val Val
85 90 95

Ser Val Cys Pro Gln Ser Leu Pro Tyr Phe Ala Ala Lys Phe Asn Leu
100 105 110

Ser Val Thr Asp Ala Ser Arg Arg Leu Cys Gly Phe Leu Lys Ser Leu
115 120 125

Gly Val His Tyr Val Phe Asp Thr Thr Ile Ala Ala Asp Phe Ser Ile
130 135 140

Leu Glu Ser Gln Lys Glu Phe Val Arg Arg Tyr Arg Gln His Ser Glu
145 150 155 160

Glu Glu Arg Thr Leu Pro Met Leu Thr Ser Ala Cys Pro Gly Trp Val
165 170 175

Arg Tyr Ala Glu Arg Val Leu Gly Arg Pro Ile Thr Ala His Leu Cys
180 185 190

Thr Ala Lys Ser Pro Gln Gln Val Met Gly Ser Leu Val Lys Asp Tyr
195 200 205

Phe Ala Arg Gln Gln Asn Leu Ser Pro Glu Lys Ile Phe His Val Ile
210 215 220

Val Ala Pro Cys Tyr Asp Lys Lys Leu Glu Ala Leu Gln Glu Ser Leu
225 230 235 240

Pro Pro Ala Leu His Gly Ser Arg Gly Ala Asp Cys Val Leu Thr Ser
245 250 255

Gly Glu Ile Ala Gln Ile Met Glu Gln Gly Asp Leu Ser Val Arg Asp
260 265 270

Ala Ala Val Asp Thr Leu Phe Gly Asp Leu Lys Glu Asp Lys Val Thr
275 280 285

Arg His Asp Gly Ala Ser Ser Asp Gly His Leu Ala His Ile Phe Arg
290 295 300

His Ala Ala Lys Glu Leu Phe Asn Glu Asp Val Glu Glu Val Thr Tyr
305 310 315 320

Arg Ala Leu Arg Asn Lys Asp Phe Gln Glu Val Thr Leu Glu Lys Asn
325 330 335

Gly Glu Val Val Leu Arg Phe Ala Ala Ala Tyr Gly Phe Arg Asn Ile
340 345 350

Gln Asn Met Ile Leu Lys Leu Lys Lys Gly Lys Phe Pro Phe His Phe
355 360 365

Val Glu Val Leu Ala Cys Ala Gly Gly Cys Leu Asn Gly Arg Gly Gln
370 375 380

Ala Gln Thr Pro Asp Gly His Ala Asp Lys Ala Leu Leu Arg Gln Met
385 390 395 400

Glu Gly Ile Tyr Ala Asp Ile Pro Val Arg Arg Pro Glu Ser Ser Ala
405 410 415

His Val Gln Glu Leu Tyr Gln Glu Trp Leu Glu Gly Ile Asn Ser Pro
420 425 430

Lys Ala Arg Glu Val Leu His Thr Thr Tyr Gln Ser Gln Glu Arg Gly
435 440 445

Thr His Ser Leu Asp Ile Lys Trp
450 455

<210> 67
<211> 408
<212> PRT
<213> Homo sapiens

<400> 67

Met Lys Cys Glu His Cys Thr Arg Lys Glu Cys Ser Lys Lys Thr Lys
1 5 10 15

Thr Asp Asp Gln Glu Asn Val Ser Ala Asp Ala Pro Ser Pro Ala Gln
20 25 30

Glu Asn Gly Glu Lys Cys Asp Thr Ser Lys His Lys Val Leu Val Val
35 40 45

Ser Val Cys Pro Gln Ser Leu Pro Tyr Phe Ala Ala Lys Phe Asn Leu
50 55 60

Ser Val Thr Asp Ala Ser Arg Arg Leu Cys Gly Phe Leu Lys Ser Leu
65 70 75 80

Gly Val His Tyr Val Phe Asp Thr Thr Ile Ala Ala Asp Phe Ser Ile
85 90 95

Leu Glu Ser Gln Lys Glu Phe Val Arg Arg Tyr Arg Gln His Ser Glu
100 105 110

Glu Glu Arg Thr Leu Pro Met Leu Thr Ser Ala Cys Pro Gly Trp Val
115 120 125

Arg Tyr Ala Glu Arg Val Leu Gly Arg Pro Ile Thr Ala His Leu Cys
130 135 140

Thr Ala Lys Ser Pro Gln Gln Val Met Gly Ser Leu Val Lys Asp Tyr
145 150 155 160

Phe Ala Arg Gln Gln Asn Leu Ser Pro Glu Lys Ile Phe His Val Ile
165 170 175

Val Ala Pro Cys Tyr Asp Lys Lys Leu Glu Ala Leu Gln Glu Ser Leu
180 185 190

Pro Pro Ala Leu His Gly Ser Arg Gly Ala Asp Cys Val Leu Thr Ser
195 200 205

Gly Glu Ile Ala Gln Ile Met Glu Gln Gly Asp Leu Ser Val Arg Asp
210 215 220

Ala Ala Val Asp Thr Leu Phe Gly Asp Leu Lys Glu Asp Lys Val Thr
225 230 235 240

```
Arg His Asp Gly Ala Ser Ser Asp Gly His Leu Ala His Ile Phe Arg
                245                 250                 255

His Ala Ala Lys Glu Leu Phe Asn Glu Asp Val Glu Glu Val Thr Tyr
            260                 265                 270

Arg Ala Leu Arg Asn Lys Asp Phe Gln Glu Val Thr Leu Glu Lys Asn
        275                 280                 285

Gly Glu Val Val Leu Arg Phe Ala Ala Ala Tyr Gly Phe Arg Asn Ile
    290                 295                 300

Gln Asn Met Ile Leu Lys Leu Lys Lys Gly Lys Phe Pro Phe His Phe
305                 310                 315                 320

Val Glu Val Leu Ala Cys Ala Gly Gly Cys Leu Asn Gly Arg Gly Gln
                325                 330                 335

Ala Gln Thr Pro Asp Gly His Ala Asp Lys Ala Leu Leu Arg Gln Met
            340                 345                 350

Glu Gly Ile Tyr Ala Asp Ile Pro Val Arg Arg Pro Glu Ser Ser Ala
        355                 360                 365

His Val Gln Glu Leu Tyr Gln Glu Trp Leu Glu Gly Ile Asn Ser Pro
    370                 375                 380

Lys Ala Arg Glu Val Leu His Thr Thr Tyr Gln Ser Gln Glu Arg Gly
385                 390                 395                 400

Thr His Ser Leu Asp Ile Lys Trp
                405
```

<210> 68
<211> 502
<212> PRT
<213> Homo sapiens

<400> 68

```
Met Lys Cys Glu His Cys Thr Arg Lys Glu Cys Ser Lys Lys Thr Lys
1               5                   10                  15

Thr Asp Asp Gln Glu Asn Val Ser Ala Asp Ala Pro Ser Pro Ala Gln
            20                  25                  30

Glu Asn Gly Glu Lys Gly Glu Phe His Lys Leu Ala Asp Ala Lys Ile
        35                  40                  45

Phe Leu Ser Asp Cys Leu Ala Cys Asp Ser Cys Met Thr Ala Glu Glu
    50                  55                  60

Gly Val Gln Leu Ser Gln Gln Asn Ala Lys Asp Phe Phe Arg Val Leu
65                  70                  75                  80

Asn Leu Asn Lys Lys Cys Asp Thr Ser Lys His Lys Val Leu Val Val
                85                  90                  95
```

Ser Val Cys Pro Gln Ser Leu Pro Tyr Phe Ala Ala Lys Phe Asn Leu
100 105 110

Ser Val Thr Asp Ala Ser Arg Arg Leu Cys Gly Phe Leu Lys Ser Leu
115 120 125

Gly Val His Tyr Val Phe Asp Thr Thr Ile Ala Ala Asp Phe Ser Ile
130 135 140

Leu Glu Ser Gln Lys Glu Phe Val Arg Arg Tyr Arg Gln His Ser Glu
145 150 155 160

Glu Glu Arg Thr Leu Pro Met Leu Thr Ser Ala Cys Pro Gly Trp Val
165 170 175

Arg Tyr Ala Glu Arg Val Leu Gly Arg Pro Ile Thr Ala His Leu Cys
180 185 190

Thr Ala Lys Ser Pro Gln Gln Val Met Gly Ser Leu Val Lys Asp Tyr
195 200 205

Phe Ala Arg Gln Gln Asn Leu Ser Pro Glu Lys Ile Phe His Val Ile
210 215 220

Val Ala Pro Cys Tyr Asp Lys Lys Leu Glu Ala Leu Gln Glu Ser Leu
225 230 235 240

Pro Pro Ala Leu His Gly Ser Arg Gly Ala Asp Cys Val Leu Thr Ser
245 250 255

Glu Ile Ser Gln Ala Trp Trp Cys Thr Pro Val Ile Thr Ala Thr Arg
260 265 270

Glu Ala Ala Ala Arg Glu Ser Leu Glu Pro Gly Arg Gln Arg Leu Gln
275 280 285

Arg Asp Lys Ile Ala Pro Leu Asp Ser Ser Leu Gly Gly Gly Gly Glu
290 295 300

Ile Ala Gln Ile Met Glu Gln Gly Asp Leu Ser Val Arg Asp Ala Ala
305 310 315 320

Val Asp Thr Leu Phe Gly Asp Leu Lys Glu Asp Lys Val Thr Arg His
325 330 335

Asp Gly Ala Ser Ser Asp Gly His Leu Ala His Ile Phe Arg His Ala
340 345 350

Ala Lys Glu Leu Phe Asn Glu Asp Val Glu Glu Val Thr Tyr Arg Ala
355 360 365

Leu Arg Asn Lys Asp Phe Gln Glu Val Thr Leu Glu Lys Asn Gly Glu
370 375 380

Val Val Leu Arg Phe Ala Ala Ala Tyr Gly Phe Arg Asn Ile Gln Asn
385 390 395 400

Met Ile Leu Lys Leu Lys Lys Gly Lys Phe Pro Phe His Phe Val Glu
405 410 415

Val Leu Ala Cys Ala Gly Gly Cys Leu Asn Gly Arg Gly Gln Ala Gln
420 425 430

Thr Pro Asp Gly His Ala Asp Lys Ala Leu Leu Arg Gln Met Glu Gly
435 440 445

Ile Tyr Ala Asp Ile Pro Val Arg Arg Pro Glu Ser Ser Ala His Val
450 455 460

Gln Glu Leu Tyr Gln Glu Trp Leu Glu Gly Ile Asn Ser Pro Lys Ala
465 470 475 480

Arg Glu Val Leu His Thr Thr Tyr Gln Ser Gln Glu Arg Gly Thr His
485 490 495

Ser Leu Asp Ile Lys Trp
500

<210> 69
<211> 448
<212> PRT
<213> Clostridium tetani

<400> 69

Met His Asn Asp Tyr Arg Glu Ile Phe Lys Arg Leu Ser Lys Ser Tyr
1 5 10 15

Tyr Asp Asp Thr Phe Glu Lys Glu Val Glu Asn Ile Leu Ser Ser His
20 25 30

Ser Met Asp Arg Glu Lys Leu Ala Lys Ile Ile Ser Ile Leu Cys Gly
35 40 45

Val Asn Ile Glu His Ser Glu Asn Tyr Ile Ser Asn Leu Lys Asn Ala
50 55 60

Ile Lys Asn Tyr Thr Ala Ser Ala Glu Lys Val Val Thr Lys Leu Pro
65 70 75 80

Cys Ser Thr Gln Cys Ala Lys Asp Gly Asp Ile Ile Cys Glu Lys Ser
85 90 95

Cys Pro Val Asn Ala Ile Phe Arg Asp Pro Asn Asp Asn Asn Ile Tyr
100 105 110

Ile Asn Asp Glu Leu Cys Leu Asp Cys Gly Leu Cys Val Arg Asn Cys
115 120 125

Pro Ser Gly Ser Ile Leu Asp Lys Lys Glu Phe Ile Pro Leu Ala Glu
130 135 140

Leu Leu Lys Ser Glu Ser Ile Val Ile Ala Ala Val Ala Pro Ala Ile
145 150 155 160

Met Gly Gln Phe Gly Glu Asn Thr Thr Ile Asn Gln Leu Arg Thr Ala
165 170 175

Phe Lys Lys Leu Gly Phe Thr Asp Met Val Glu Val Ala Phe Phe Ala
180 185 190

Asp Met Leu Thr Leu Lys Glu Ala Val Glu Tyr Asp His Phe Val Lys
195 200 205

Asp Glu Gln Asp Phe Met Ile Thr Ser Cys Cys Cys Pro Met Trp Val
210 215 220

Gly Met Leu Lys Lys Val Tyr Asn Asp Leu Val Lys Tyr Val Ser Pro
225 230 235 240

Ser Val Ser Pro Met Ile Ala Ala Gly Arg Val Leu Lys Leu Leu Asn
245 250 255

Pro Asn Cys Lys Val Val Phe Val Gly Pro Cys Ile Ala Lys Lys Ala
260 265 270

Glu Ala Arg Glu Lys Asp Leu Leu Gly Asp Ile Asp Phe Val Leu Thr
275 280 285

Phe Thr Glu Leu Arg Asp Ile Phe Asp Val Phe Asp Ile Gln Pro Glu
290 295 300

Asn Leu Glu Glu Asp Phe Ser Ser Glu Tyr Ala Ser Lys Gly Gly Arg
305 310 315 320

Leu Tyr Ala Arg Thr Gly Gly Val Ser Ile Ala Val Ser Glu Ala Ile
325 330 335

Glu Lys Leu Phe Pro Asn Lys Tyr Lys Phe Leu Lys Thr Ile Gln Ala
340 345 350

Asp Gly Val Lys Gly Cys Lys Ser Leu Leu Asp Lys Ile Lys Gln Glu
355 360 365

Asp Ile Ser Ala Asn Phe Val Glu Gly Met Gly Cys Val Gly Gly Cys
370 375 380

Val Gly Gly Pro Lys Val Ile Ile Asp Pro Ser Glu Gly Arg Asn Ala
385 390 395 400

Val Asn Asn Phe Ala Glu Asn Ser Ser Ile Lys Val Ser Val Asp Ser
405 410 415

Asn Cys Met Asn Asp Ile Leu Ser Lys Ile Asn Ile Asn Ser Val Glu
420 425 430

Asp Phe Lys Asp Lys Asp Lys Ile Ser Ile Phe Glu Arg Glu Phe Lys
435 440 445

<210> 70

<211> 459
<212> PRT
<213> Desulfovibrio desulfuricans

<400> 70

Met Tyr Phe Arg Thr Tyr Asp Asn Thr Ile Asn Phe Glu Ile Met Val
1 5 10 15

Arg Ile Ala Lys Ala Phe His Gly Asp Ser Phe Glu Glu Gln Val Ala
20 25 30

Arg Ile Pro Leu Glu Met Arg Pro Arg Lys Ala His Ser Ser Arg Cys
35 40 45

Cys Ile Tyr Arg Asp Arg Ala Ile Ile Arg Tyr Arg Cys Met Ala Met
50 55 60

Leu Gly Tyr Ala Ile Glu Asp Glu Thr Asp Glu Leu Thr Ser Leu Ser
65 70 75 80

Gln Tyr Ala Lys Gly Ala Leu Glu Arg Asp Ser Ile Gln Gly Ser Met
85 90 95

Leu Thr Phe Ile Asp Glu Ala Cys Asn Gly Cys Val Arg Thr His Tyr
100 105 110

Glu Ala Thr Ser Ala Cys Arg Gly Cys Leu Ala Glu Ala Cys Val Gln
115 120 125

His Cys Pro Lys Asp Ala Val Arg Ile Val Asp Gly Lys Ser Arg Ile
130 135 140

Asp Pro Asp Lys Cys Val Gln Cys Gly Lys Cys Met Asn Val Cys Pro
145 150 155 160

Tyr His Ala Ile Val Gln Ile Pro Ile Pro Cys Glu Glu Ser Cys Pro
165 170 175

Thr Gly Ala Ile Ser Lys Asp Glu Cys Gly Lys Gln Val Ile Asp Tyr
180 185 190

Asp Arg Cys Ile Phe Cys Gly Lys Cys Met Ala Ala Cys Pro Phe Ala
195 200 205

Ala Val Leu Glu Lys Ser Gln Met Ile Asp Val Leu Arg Arg Ile Arg
210 215 220

Glu Gly Arg Lys Val Val Ala Ile Val Ala Pro Ala Ile Ala Gly Gln
225 230 235 240

Val Gln Ala Pro Met Ser Arg Leu Ala Thr Ala Leu Arg Gln Leu Gly
245 250 255

Phe Ala Asp Val Ala Glu Val Ala Ser Gly Ala Asp Thr Thr Ala Arg
260 265 270

Leu Glu Ala Asp Glu Phe Val Glu Arg Met Glu His Gly Ala Ala Phe

275 280 285

Met Thr Ser Ser Cys Cys Pro Ala Tyr Thr Gln Leu Val Asp Lys His
290 295 300

Leu Pro Glu Leu Ala Pro Phe Val Ser Asp Thr Arg Thr Pro Met His
305 310 315 320

Tyr Thr Ala Ala Met Val Lys Asp His Asp Pro Asp Met Val Thr Val
325 330 335

Phe Ile Gly Pro Cys Val Ala Lys Arg Asn Glu Gly Lys His Asp Glu
340 345 350

Leu Val Asp His Val Leu Thr Phe Gln Glu Met Val Ala Met Leu Thr
355 360 365

Ala Ala Gly Ile Ser Val Asp Ala Cys Glu Asp Gly Arg Phe Met Phe
370 375 380

Pro Ala Met Arg Glu Gly Arg Ser Phe Pro Val Ser Gly Gly Val Thr
385 390 395 400

Ala Gly Val Gln Ala His Ile Gly Thr Arg Ala Glu Val Arg Pro Leu
405 410 415

Ser Val Asp Gly Leu Asn Lys Lys Thr Phe Arg Gln Leu Lys Thr Trp
420 425 430

Ala Lys Lys Gly Cys Glu Gly Asn Phe Val Glu Val Met Gly Cys Gln
435 440 445

Gly Gly Cys Val Ala Gly Pro Ala Ile Val Met
450 455

<210> 71
<211> 494
<212> PRT
<213> Clostridium tetani

<400> 71

Met Ile Val Phe Glu Asn Gln Leu Lys Lys Leu Lys Tyr Leu Val Leu
1 5 10 15

Lys Glu Val Ala Lys Met Thr Leu Glu Asp Arg Leu Gly Glu Glu Asp
20 25 30

Ile Glu Arg Ile Ser Phe Asp Ile Ile Lys Gly Asp Lys Ala Glu Tyr
35 40 45

Arg Cys Cys Val Tyr Lys Glu Arg Ala Ile Val Tyr Glu Arg Ala Lys
50 55 60

Leu Ala Thr Gly Cys Leu Pro Asn Gly Gln Val Ala Glu Glu Phe Val
65 70 75 80

His Val Glu Asp Asp Asp Gln Ile Ile Tyr Val Ile Asp Ala Ala Cys

Asp Lys Cys Pro Ile Asn Lys Tyr Val Val Thr Glu Ala Cys Arg Gly
100 105 110

Cys Leu Gln His Lys Cys Met Glu Val Cys Pro Ala Gly Ser Ile Asn
115 120 125

Arg Ala Ala Gly Lys Ala Tyr Ile Asn His Glu Thr Cys Lys Glu Cys
130 135 140

Gly Leu Cys Glu Ser Ala Cys Pro Tyr Asn Ala Ile Ala Glu Val Met
145 150 155 160

Arg Pro Cys Arg Arg Ala Cys Pro Thr Gly Ala Leu Gln Met Asn Leu
165 170 175

Glu Asp Asn Lys Ala Thr Ile Asn Lys Glu Asp Cys Ile Asn Cys Gly
180 185 190

Ser Cys Met Ser Val Cys Pro Phe Gly Ala Ile Ser Asp Lys Ser Tyr
195 200 205

Ile Val Asp Ile Thr Lys Ala Leu Lys Asn Asn Lys Lys Val Tyr Ala
210 215 220

Met Val Ala Pro Ala Ile Thr Gly Gln Phe Gly Lys Asp Val Ser Val
225 230 235 240

Gly Lys Met Lys Asn Ala Phe Lys Ala Met Gly Phe Glu Asp Met Leu
245 250 255

Glu Val Ala Cys Gly Ala Asp Ala Val Ala Ala His Glu Ser Glu Glu
260 265 270

Phe Ile Glu Arg Leu Glu Ser Gly Lys Lys Tyr Met Thr Thr Ser Cys
275 280 285

Cys Pro Gly Phe Leu Gly Tyr Ile Glu Lys Lys Phe Pro Asp Gln Leu
290 295 300

Glu Asn Val Ser Asn Thr Val Ser Pro Met Val Ala Ile Gly Arg Met
305 310 315 320

Ile Lys Lys Glu Tyr Glu Asp Ser Val Val Val Phe Val Gly Pro Cys
325 330 335

Thr Ala Lys Lys Ala Glu Ile Lys Arg Lys Gly Ile Lys Asp Ala Val
340 345 350

Asp Tyr Val Met Thr Phe Glu Glu Ile Ala Ala Leu Met Gly Ala Phe
355 360 365

Glu Ile Asp Pro Ala Glu Cys Glu Glu Glu Asp Ile Asn Asp Gly Ser
370 375 380

Asn Tyr Gly Arg Gly Phe Ala Gln Gly Gly Gly Val Val Ser Ala Ile
385 390 395 400

Gln Asn Cys Ile Lys Asp Lys Glu Gly Ile Lys Phe Asn Pro Leu Arg
405 410 415

Val Ser Gly Pro Asp Gln Ile Lys Arg Ala Met Ile Met Ala Lys Val
420 425 430

Gly Lys Leu Ser Glu Asn Phe Ile Glu Gly Met Met Cys Glu Gly Gly
435 440 445

Cys Ile Gly Gly Pro Ala Thr Met Val Ser Ala Val Lys Ala Lys Ala
450 455 460

Pro Leu Met Lys Phe Ser Lys Ser Ser Thr Ile Lys Asp Val Lys Asp
465 470 475 480

Asn Glu Val Leu Asp Lys Tyr Lys Asp Ile Asn Met Glu Arg
485 490

<210> 72
<211> 203
<212> PRT
<213> Arabidopsis thaliana

<400> 72

Met Asp Leu Ile Lys Leu Lys Gly Val Asp Phe Lys Asp Leu Glu Glu
1 5 10 15

Ser Pro Leu Asp Arg Val Leu Thr Asn Val Thr Glu Glu Gly Asp Leu
20 25 30

Tyr Gly Val Ala Gly Ser Ser Gly Gly Tyr Ala Glu Thr Ile Phe Arg
35 40 45

His Ala Ala Lys Ala Leu Phe Gly Gln Thr Ile Glu Gly Pro Leu Glu
50 55 60

Phe Lys Thr Leu Arg Asn Ser Asp Phe Arg Glu Val Thr Leu Gln Leu
65 70 75 80

Glu Gly Lys Thr Val Leu Lys Phe Ala Leu Cys Tyr Gly Phe Gln Asn
85 90 95

Leu Gln Asn Ile Val Arg Arg Val Lys Thr Arg Lys Cys Asp Tyr Gln
100 105 110

Tyr Val Glu Ile Met Ala Cys Pro Ala Gly Cys Leu Asn Gly Gly Gly
115 120 125

Gln Ile Lys Pro Lys Thr Gly Gln Ser Gln Lys Glu Leu Ile His Ser
130 135 140

Leu Glu Ala Thr Tyr Met Asn Asp Thr Thr Leu Asn Thr Asp Pro Tyr
145 150 155 160

Gln Asn Pro Thr Ala Lys Arg Leu Phe Glu Glu Trp Leu Lys Glu Pro
165 170 175

Gly Ser Asn Glu Ala Lys Lys Tyr Leu His Thr Gln Tyr His Pro Val
180 185 190

Val Lys Ser Val Thr Ser Gln Leu Asn Asn Trp
195 200

<210> 73
<211> 449
<212> PRT
<213> Clostridium perfringens

<400> 73

Met Asn Lys Lys Tyr Asn Ser Leu Phe Lys Glu Leu Ile Ser Ser Tyr
1 5 10 15

Tyr Ser Glu Asp Asn Phe Asp Glu Lys Leu Asn Asp Ile Val Lys Asn
20 25 30

Asn Phe Asn Ser Lys Glu Asp Ala Ile Glu Val Leu Ser Ser Leu Cys
35 40 45

Gly Val Asp Ile Asp Lys Asn Ser Asp Asn Ile Ala Tyr Asp Ile Arg
50 55 60

Lys Ala Ile Thr Thr His Lys Ile Lys Lys Asn Ile Val Asp Lys Val
65 70 75 80

Ser Val Cys Thr Lys Asn Cys Ser Lys Glu Ser Lys Gly Lys Cys Gln
85 90 95

Ser Leu Cys Pro Phe Asp Ala Ile Leu Thr Asp Pro Ile Asp Asn Ser
100 105 110

Lys Tyr Ile Asp Pro Asn Leu Cys Gln Asn Cys Gly Ile Cys Val Gln
115 120 125

Val Cys Glu Ser Gly His Phe Leu Asp Arg Ile Glu Leu Leu Pro Ile
130 135 140

Ile Asp Leu Ile Lys Asn Asn Glu Thr Val Ile Ala Ala Val Ala Pro
145 150 155 160

Ala Ile Ala Gly Gln Phe Gly Glu Asn Val Ser Leu Asp Met Leu Arg
165 170 175

Glu Ala Phe Ile Lys Ile Gly Phe Ser Asp Met Ile Glu Val Ala Phe
180 185 190

Ala Ala Asp Met Leu Ser Ile Lys Glu Ala Val Glu Phe Asn His His
195 200 205

Val Glu Lys Thr Gly Asp Ile Leu Ile Thr Ser Cys Cys Cys Pro Met
210 215 220

Trp Val Ala Met Leu Arg Lys Cys Tyr Lys Asp Leu Val Lys Asp Val
225 230 235 240

Ser Pro Ser Val Ser Pro Met Ile Ala Ala Gly Arg Val Ile Lys Lys
245 250 255

Leu Asn Lys Asp Ala Lys Val Val Phe Ile Gly Pro Cys Ile Ala Lys
260 265 270

Lys Ala Glu Ala Arg Glu Lys Asp Leu Val Gly Ala Ile Asp Tyr Val
275 280 285

Leu Thr Phe Glu Glu Leu Asn Gly Ile Phe Glu Ala Leu Lys Ile Asp
290 295 300

Pro Ser Ser Met Lys Gly Val Pro Ser Ile Glu Tyr Thr Ser Arg Gly
305 310 315 320

Gly Arg Leu Tyr Ala Arg Thr Gly Gly Val Ser Glu Ala Ile Asn Asp
325 330 335

Val Val Lys Glu Leu Tyr Pro Asp Lys Ala Lys Ile Phe Lys Ala Val
340 345 350

Gln Ala Asn Gly Val Lys Glu Cys Lys Glu Leu Leu Asn Lys Val Gln
355 360 365

Ser Gly Glu Leu Lys Ala Asn Phe Ile Glu Gly Met Gly Cys Val Gly
370 375 380

Gly Cys Val Gly Gly Pro Lys Arg Ile Val Asp Pro Ser Ile Gly Lys
385 390 395 400

Lys His Val Asp Glu Val Ala Tyr Asn Ser Pro Ile Lys Val Ala Thr
405 410 415

His Ser His Thr Met Asp Glu Val Leu Leu Arg Leu Gly Ile Asn Ser
420 425 430

Leu Lys Ser Phe Glu Asp Lys Glu Lys Ile Ser Ile Phe Glu Arg Glu
435 440 445

Phe

<210> 74
<211> 359
<212> PRT
<213> Desulfitobacterium hafniense

<400> 74

Met Ala Gln Ser Glu Ile Met Lys Ile Arg Arg Gln Val Leu Lys Ser
1 5 10 15

Ala Leu Asp Trp Val Ser His Asp Gln Asn Arg Lys Asp Arg Ala Thr
20 25 30

Leu Ala Arg Gln Ile Ile Pro Asp Gly Thr Pro Arg Tyr Arg Cys Cys
35 40 45

Ile His Lys Glu Arg Ala Val Ile Glu Glu Arg Leu Lys Ala Val Leu
50 55 60

Glu Pro Asp Glu Gly Pro Ile Val Arg Val Leu Lys Glu Gly Cys Asn
65 70 75 80

Gly Cys Glu Met His Arg Tyr Ser Val Thr Asp His Cys Gln Asn Cys
85 90 95

Val Gly His Phe Cys Phe Thr Asn Cys Pro Lys Lys Ala Ile Leu Phe
100 105 110

Ile Asn Asn Lys Ala Phe Ile Asp Gln Thr Arg Cys Val Glu Cys Gly
115 120 125

Leu Cys Ala Arg Asn Cys Pro Tyr His Ala Ile Ile Glu Tyr Arg Arg
130 135 140

Pro Cys Glu Asp Ser Cys Pro Thr Lys Ala Ile Ser Val Arg Glu Asp
145 150 155 160

Arg Ile Ala Ser Ile Ala Glu Ala His Cys Thr Ser Cys Gly Lys Cys
165 170 175

Ile Ile Ser Cys Pro Phe Gly Ala Val Ala Glu Ser Ser Gln Leu Ile
180 185 190

His Leu Phe Glu Ala Val Arg Asn Pro Glu His Lys Ile Tyr Ala Val
195 200 205

Ile Ala Pro Ala Phe Val Gly Gln Phe Gly Arg Lys Val Ser Pro Gly
210 215 220

Gln Val Lys Ser Ala Leu Leu Lys Leu Gly Phe Gln Asp Val Leu Glu
225 230 235 240

Ala Ala Leu Gly Ala Asp Arg Thr Ile Glu Leu Glu Ala Arg Glu Tyr
245 250 255

Asp Glu Arg Leu Ala His Gly Glu Glu Phe Met Thr Ser Ser Cys Cys
260 265 270

Pro Ala Tyr Val Ser Ala Val Ile Lys Glu Lys Pro Asp Leu Phe His
275 280 285

His Ile Ser Ser Thr Leu Ser Pro Met Ala Gln Val Ala His Ile Leu
290 295 300

Lys Glu Lys Asp Pro Glu Ala Lys Ile Ala Phe Ile Gly Pro Cys Val
305 310 315 320

Ala Lys Lys Glu Glu Gly Lys Arg Pro Glu Thr Lys Val Asp Phe Val
325 330 335

Leu Thr Phe Glu Glu Leu Met Val Trp Leu Asp Tyr Ala Gly Ile Asn
340 345 350

Pro Ala Glu Glu Ser Glu Gln
355

<210> 75
<211> 790
<212> PRT
<213> Geobacter metallireducens

<400> 75

Met Cys His Trp Leu His Arg Glu Ala Gly Leu Val Tyr Asp Pro Ala
1 5 10 15

Val Asp Gln Ala Ile Asn Arg Val Ser Gly Leu Thr Leu Ser Ala Gly
20 25 30

Arg Thr Met Glu Pro Ile Ile Thr Val Lys Glu Lys Cys Arg Lys Cys
35 40 45

Tyr Cys Cys Val Arg Ser Cys Pro Val Lys Ala Ile Lys Val Ala Lys
50 55 60

Ser Tyr Thr Glu Ile Ile Val Asp Arg Cys Ile Gly Cys Gly Asn Cys
65 70 75 80

Leu Ser Asn Cys Pro Gln Gln Ala Lys Met Val Ala Asp Lys Val Gly
85 90 95

Val Thr Glu Lys Leu Leu Ser Ser Gly Glu Glu Val Ile Ala Val Leu
100 105 110

Gly Ser Ser Phe Pro Ala Phe Phe His Asn Val Thr Pro Gly Gln Leu
115 120 125

Val Ala Gly Leu Arg Lys Ile Gly Phe Ala Glu Val His Glu Gly Ser
130 135 140

Tyr Gly Ala Glu Leu Ile Ala Asp Asp Tyr Ala Arg Ile Thr Ser Glu
145 150 155 160

Lys Gly His Pro Arg Ile Ser Ser His Cys Pro Ala Ile Val Asp Leu
165 170 175

Ile Glu Arg His Tyr Pro Lys Leu Val Gly Asn Leu Val Pro Val Val
180 185 190

Ser Pro Met Val Ala Met Gly Arg Tyr Leu Lys Gly Thr Leu Gly Gln
195 200 205

His Val Arg Val Val Tyr Ile Ser Ser Cys Val Ala Asn Lys Leu Glu
210 215 220

Thr Gln Thr Gln Glu Thr Arg Gly Ala Val Asp Ile Val Leu Thr Tyr
225 230 235 240

Arg Glu Leu Glu Gly Ile Phe Arg Ser Arg Gln Ile Ala Leu Pro Ala
245 250 255

Leu Ala Asp Glu Pro Leu Asp Gly Ile Arg Pro Gly Ala Gly Arg Leu
260 265 270

Phe Pro Ile Ala Asp Gly Thr Phe Arg Ala Phe Gly Ile Pro Phe Asp
275 280 285

Pro Leu Asp Thr Glu Ile Val Ala Ala Cys Gly Glu Val Asn Val Met
290 295 300

Gly Ile Ile Asn Asp Leu Ala Ala Gly Arg Ile Ser Pro Arg Ile Ala
305 310 315 320

Asp Leu Arg Phe Cys Tyr Asp Gly Cys Ile Gly Gly Pro Gly Arg Asn
325 330 335

Arg Ala Leu Thr Glu Phe Tyr Arg Arg Asn Arg Val Ile Ala His Phe
340 345 350

Lys Gln Glu Val Pro Cys Arg Thr Val Pro Asn Ser Leu Leu Glu Ala
355 360 365

Gly Arg Val Ser Phe Gly Arg Ser Phe Ala Ser Lys Tyr Ala Lys Leu
370 375 380

Glu Ala Pro Lys Ala Asn Asp Val Arg Lys Ile Leu Asn Ala Thr Asn
385 390 395 400

Lys Phe Thr Val Lys Asp Glu Leu Asn Cys Arg Ala Cys Gly Tyr Arg
405 410 415

Thr Cys Arg Glu Tyr Ala Val Ala Val Phe Gln Gly Leu Ala Glu Ile
420 425 430

Glu Met Cys Leu Pro Tyr Asn Leu Gln Gln Leu Glu Glu Asp Arg Gly
435 440 445

Arg Leu Ile Gln Lys Tyr Glu Leu Ala Arg Arg Glu Leu Glu Arg Glu
450 455 460

Tyr Gly Asp Glu Phe Ile Val Gly Asn Asp Arg Lys Thr Leu Asp Val
465 470 475 480

Leu Gly Leu Ile Lys Gln Val Gly Pro Thr Pro Thr Thr Val Leu Ile
485 490 495

Arg Gly Glu Ser Gly Thr Gly Lys Glu Leu Thr Ala Arg Ala Ile His
500 505 510

Arg Tyr Ser Lys Arg Asn Asp Lys Pro Leu Val Thr Val Asn Cys Thr
515 520 525

Thr Ile Thr Asp Ser Leu Leu Glu Ser Glu Leu Phe Gly His Lys Arg
530 535 540

Gly Ala Phe Thr Gly Ala Val Ala Asp Lys Lys Gly Leu Phe Glu Ala
545 550 555 560

Ala Asp Gly Gly Thr Ile Phe Leu Asp Glu Ile Gly Asp Ile Thr Pro
565 570 575

Lys Leu Gln Ala Glu Leu Leu Arg Val Leu Asp Met Gly Glu Val Arg
580 585 590

Pro Val Gly Gly Thr Ala Ala Lys Lys Val Asp Val Arg Leu Ile Ala
595 600 605

Ala Thr Asn Lys Asn Leu Glu Gln Gly Val Arg Glu Gly Trp Phe Arg
610 615 620

Glu Asp Leu Tyr Tyr Arg Leu Asn Val Phe Thr Ile Thr Met Pro Pro
625 630 635 640

Leu Arg Ser Arg Val Glu Ser Ile Pro Ile Leu Val His His Phe Met
645 650 655

Asp Lys Ala Ser Thr Lys Leu Asn Lys Arg Met Val Gly Ile Glu Asp
660 665 670

Arg Ala Val Lys Ala Leu Thr Lys Tyr Pro Trp Pro Gly Asn Ile Arg
675 680 685

Glu Met Gln Asn Val Ile Glu Arg Ala Ala Val Leu Thr His Asp Gly
690 695 700

Val Ile Arg Val Glu Asn Phe Pro Leu Ala Leu Ser Glu Gly Leu Glu
705 710 715 720

Glu Gly Phe Ala Thr Gly Leu Asp Ile His Ala Ala Ser Phe Arg Ser
725 730 735

Glu Arg Glu Gln His Met Gly Lys Leu Glu Lys Lys Leu Ile Gln Arg
740 745 750

Tyr Leu Thr Glu Ala Asn Gly Asn Ile Ser Arg Ala Ala Lys Leu Ala
755 760 765

Asn Ile Pro Arg Arg Thr Phe Tyr Arg Leu Leu Asp Lys Tyr Arg Leu
770 775 780

Arg Glu Arg Asp Val Arg
785 790

<210> 76
<211> 450
<212> PRT
<213> Clostridium acetobutylicum

<400> 76

Met Asn Asn Lys Tyr Ile Glu Leu Phe Lys Ser Leu Val Asp Ser Tyr
1 5 10 15

Tyr Asn Asp Thr Phe Asp Ser Phe Val Tyr His Ile Leu Ser Asp Glu
20 25 30

Glu Val Asp Lys Lys Glu Leu Ser Lys Val Ile Ser Ser Leu Cys Gly
35 40 45

Val Ser Val Glu Phe Lys Asp Thr Glu Thr Tyr Ile Ser Glu Leu Lys
50 55 60

Lys Ala Ile Ser Asn Tyr Lys Cys Thr Asp Asn Ile Val Glu Lys Ile
65 70 75 80

Lys Glu Cys Asp Ser Ser Cys His Ser Asn Glu Gly Glu Thr Pro Cys
85 90 95

Gln Lys Ser Cys Pro Phe Asp Ala Ile Leu Val Asp Lys Asn Thr Lys
100 105 110

Thr Ser His Ile Gln Lys Asp Leu Cys Thr Asp Cys Gly Asn Cys Ile
115 120 125

Thr Ser Cys Pro Ser Gly Ser Ile Leu Asp Lys Ile Glu Phe Met Pro
130 135 140

Leu Leu Asn Leu Phe Lys Asn Asn Glu Thr Val Ile Ala Ala Val Ala
145 150 155 160

Pro Ala Ile Ala Gly Gln Phe Gly Glu Asn Val Ser Leu Glu Met Leu
165 170 175

Arg Thr Ala Phe Lys Lys Val Gly Phe Ala Asp Met Val Glu Val Ala
180 185 190

Phe Phe Ala Asp Met Leu Thr Ile Lys Glu Ala Phe Glu Phe Asn Glu
195 200 205

Leu Val Asn Ser Lys Asp Asp Leu Met Ile Thr Ser Cys Cys Cys Pro
210 215 220

Met Trp Val Ser Met Ile Arg Lys Ile Tyr Lys Asp Leu Ala Arg His
225 230 235 240

Val Ser Pro Ser Val Ser Pro Met Ile Ala Ser Gly Arg Val Ile Lys
245 250 255

Lys Leu Asn Pro Asn Cys Lys Val Val Phe Ile Gly Pro Cys Ile Ala
260 265 270

Lys Lys Ala Glu Ser Arg Ser Gln Asp Ile Ser Asp Ala Ile Asp Phe
275 280 285

Val Leu Thr Phe Glu Glu Leu Lys Gly Ile Phe Asp Val Leu Asp Ile
290 295 300

Asp Pro Glu Lys Leu Pro Glu Thr His Thr Lys Ser Tyr Ala Ser Arg

305 310 315 320

Glu Gly Arg Leu Tyr Gly Arg Thr Gly Gly Val Ser Thr Ser Val Asp
325 330 335

Glu Ala Val Lys Arg Ile Phe Pro Asn Lys His His Leu Phe Lys Ser
340 345 350

Thr Lys Val Asp Gly Val Lys Asp Cys Lys Asp Ile Leu Asn Lys Thr
355 360 365

Gln Ala Gly Asn Ile Gly Ala Asn Phe Leu Glu Gly Met Gly Cys Val
370 375 380

Gly Gly Cys Val Gly Gly Pro Lys Ala Ile Val His Lys Asp Gln Gly
385 390 395 400

Arg Glu Ser Val Asn Lys Thr Ala Glu Ser Ser Glu Ile Lys Ile Ser
405 410 415

Val Asp Ser Glu Arg Met Lys Asp Ile Leu Ser Arg Ile Gly Ile Asn
420 425 430

Ser Ile Glu Asp Phe Gly Asp Lys Ser Lys Val Asp Ile Phe Glu Arg
435 440 445

Arg Phe
450

<210> 77
<211> 106
<212> PRT
<213> Shewanella oneidensis

<400> 77

Met Asn Lys Lys Lys His Leu Phe Ala Glu Asp Ser Phe Phe Leu Ser
1 5 10 15

Arg Arg Lys Phe Met Ala Val Gly Ala Ala Phe Val Ala Ala Leu Ala
20 25 30

Ile Pro Ile Gly Trp Phe Thr Ser Lys Leu Glu Arg Arg Asn Glu Tyr
35 40 45

Ile Lys Ala Arg Ser Gln Gly Leu Tyr Lys Asp Asp Ser Leu Ala Lys
50 55 60

Thr Arg Val Ser His Ala Asn Pro Ala Val Glu Lys Tyr Tyr Lys Glu
65 70 75 80

Phe Gly Gly Glu Pro Leu Gly His Met Ser His Glu Leu Leu His Thr
85 90 95

His Phe Val Asp Arg Thr Lys Leu Ser Ser
100 105

<210> 78

<211> 504
<212> PRT
<213> Entamoeba histolytica

<400> 78

Met Ser Thr Gln Leu Thr Pro Leu Arg Asn Lys Ile Ile Ser Glu Val
1 5 10 15

Val Lys Cys Phe Lys Ser Gly Arg Phe Ile Glu Asp Ile Asp Lys Leu
20 25 30

Pro Thr Ile Leu Thr Asp Gly Asp Gly Trp Lys Pro Thr Ser Lys Phe
35 40 45

Val His Ser Arg Glu Gln Glu Glu Gly Ile Tyr Arg Glu Lys Val Leu
50 55 60

Ser Val Leu Gly Phe Val Asp Gly Glu Tyr Asp Asp Ile Thr Pro Leu
65 70 75 80

His Val Tyr Ala Gln Lys Ala Leu Glu Arg Thr Ser Leu His Glu Pro
85 90 95

Val Phe Gly Ile Ser Gln Lys Gly Cys Asn Lys Cys His Phe Asn Gly
100 105 110

Tyr Phe Val Thr Gln Ala Cys Glu Gly Cys Thr Ser Arg Pro Cys Ser
115 120 125

Val Asn Cys Pro Lys Lys Cys Ile Ser Phe Gly Glu Asp Gly Arg Ala
130 135 140

Val Ile Asn Gln Asn Asn Cys Ile Lys Cys Gly Arg Cys Tyr Lys Phe
145 150 155 160

Cys Pro Tyr Gly Ala Ile Ile Ser Lys Ser Val Pro Cys Val Lys Ala
165 170 175

Cys Pro Cys Gly Ala Met Leu Asp Ser Pro Glu Gly Val Lys Thr Ile
180 185 190

Asp Phe Glu Lys Cys Ile Asn Cys Gly Gly Cys Met Arg Ala Cys Pro
195 200 205

Phe Gly Ala Ile Leu Pro Arg Ser Asn Leu Ile Asp Val Leu Lys Ile
210 215 220

Leu Pro Thr Lys Lys Val Val Ala Cys Pro Ala Pro Ser Ile Ala Ala
225 230 235 240

His Phe Gly Lys Tyr Asp Leu Ala Leu Val Ser Gly Gly Leu Ile Gln
245 250 255

Val Gly Phe Thr Ser Val Glu Asp Val Ser Tyr Gly Ala Asp Leu Cys
260 265 270

Ala Leu Asn Glu Ala Lys Glu Phe Glu Glu Arg Ile Val Lys Asn Lys

275 280 285

Lys Asp Phe Met Thr Thr Ser Cys Cys Pro Ala Tyr Ile Asn Ala Ile
290 295 300

Asn Lys His Met Pro Glu Leu Lys Glu Asn Val Ser His Thr Pro Thr
305 310 315 320

Pro Met His Phe Ala Thr Gln Ala Val Lys Asp Arg Asp Gln Glu Thr
325 330 335

Val Thr Val Phe Ile Gly Pro Cys Asn Ala Lys Arg Trp Glu Thr Leu
340 345 350

Gln Asp Ser Thr Thr Asp Tyr Cys Leu Thr Phe Asp Glu Ile Phe Gly
355 360 365

Leu Phe Glu Gly Ser Gly Ile Asp Leu Ser Lys Val Gln Pro Tyr Thr
370 375 380

Phe Val Asp Lys Ala His Lys Glu Gly Lys Ile Phe Ala Val Ser Gly
385 390 395 400

Gly Val Ala Ser Ala Val Ala Ser Leu Leu Pro Lys Glu Val Pro Asp
405 410 415

Gly Val Ile Lys Pro Thr Ile Ile Asp Gly Phe Ser Gln Glu Asn Phe
420 425 430

Lys Arg Leu Lys Asn Phe Lys Lys Asn Ile Thr Gly Asn Leu Val Glu
435 440 445

Val Met Val Cys Glu Gly Gly Cys Ala Tyr Gly Pro Gly Cys Pro Gly
450 455 460

Leu Asn Thr Pro Ala Thr Ser Ala Lys Ile Lys Ile Ala Val Asp Lys
465 470 475 480

Met Glu Ala His Pro Glu Gly Arg Trp Val Gly Leu Pro Asn Ser Gln
485 490 495

Ile Lys Pro Ile Lys Val Glu Asn
500

<210> 79
<211> 560
<212> PRT
<213> Cryptosporidium parvum

<400> 79

Met Phe Ser Thr Ala Val Lys Leu Ala Asn Leu Asp Asp Tyr Leu Glu
1 5 10 15

Ser Ser Gln Asp Cys Ile Val Ser Leu Leu Ser Asp Lys Asp Asp Thr
20 25 30

Lys Pro Lys Ile Ala Val Met Arg Pro Ala Lys Ala Gln Gly Asn Lys

35/Closed

40 45

Asp Asp Lys Lys Ser Gly Thr Ser Asp Lys Ala Thr Val Asn Val Ala
50 55 60

Asp Cys Leu Ala Cys Ser Gly Cys Val Thr Ser Ala Glu Ala Lys Leu
65 70 75 80

Leu Glu Asp Gln Asn Val Ser Glu Phe Met Asn Ile Leu Lys Gln Lys
85 90 95

Arg Leu Thr Val Val Ser Ile Ser Asn Gln Ser Cys Ser Ser Phe Ala
100 105 110

Cys His Leu Asn Cys Asp Leu Ile Thr Ile Gln Arg Lys Leu Ser Gly
115 120 125

Leu Phe Lys His Ile Gly Ala Arg Phe Val Met Asn Ser Thr Ile Ser
130 135 140

Glu Tyr Ile Ser Leu Leu Glu Thr Lys Tyr Glu Phe Ile Ser Arg Tyr
145 150 155 160

Lys Ala Lys Ser Asp Leu Pro Met Ile Ile Ser His Cys Pro Gly Trp
165 170 175

Ile Cys Tyr Ser Glu Lys Ser Leu Asn Ser Ser Val Leu Pro Leu Leu
180 185 190

Ser Lys Val Arg Ser Ala Gln Gln Leu Gln Gly Ile Leu Ile Lys Thr
195 200 205

Leu Thr Leu Glu Ile Tyr Asn Gln Leu Leu Phe Leu Tyr Lys Phe Arg
210 215 220

Leu Ser Asn Ser Tyr Arg Thr Asn Met Asn Val Lys Ser Thr Phe Thr
225 230 235 240

Gln Asn Asp Asn Phe Val Glu Gln Ser Asp Ile Phe His Val Ala Ile
245 250 255

Met Pro Cys His Asp Lys Lys Leu Glu Ser Thr Arg Ser Ser Leu Ser
260 265 270

Leu Lys Ser Ser Asp Lys Asn Ser Ser Cys Pro Glu Val Asp Ile Val
275 280 285

Leu Ala Thr Ser Glu Val Gly Glu Ile Ile Lys Leu Ala Gly Phe Asn
290 295 300

Ser Leu Leu Asp Val Pro Glu Ala Pro Leu Asp Asn Leu Trp Leu Asn
305 310 315 320

Gln Asn Phe Gln Ile Thr Lys Lys His Asn Leu Ser Leu Leu Ile Thr
325 330 335

Glu Asn Tyr Val Ser Asn Gln Ile Leu Asn Gln Phe Ser Trp Leu Ile
340 345 350

Pro Ser Tyr Phe Asn Ser Asn Ser Gly Gly Phe Cys Glu Tyr Ile Ile
355 360 365

Arg Ser Ala Ile Lys Glu Leu Ala Gly Asp His Ile Asp Asn Lys Val
370 375 380

Gln Leu Pro Phe Asn Lys Leu Lys Asn Asp Ile Leu Glu Ala Lys Tyr
385 390 395 400

Ile Lys Asn Asn Val Glu Leu Asn Tyr Cys Leu Ala Tyr Gly Phe Arg
405 410 415

Ala Ile Gln Ser Ile Ser Arg Lys Leu Asn Leu Gln Lys Asn Ala Ser
420 425 430

Gln Asn Thr Gln Tyr Lys Gln Ser Val Val Asn His Val Asn Tyr His
435 440 445

Leu Ile Glu Ala Met Ala Cys Pro Thr Gly Cys Val Ser Gly Gly Gly
450 455 460

Gln Ile Leu Ser Gln Asn Asp Gln Asn Asp Asp Asn Ser Asp Leu Asn
465 470 475 480

Lys Leu Arg Lys Asn Ile Lys Phe Ile Asp Glu Val Gln Glu Ala Leu
485 490 495

Tyr Lys Gly Ile Asn Leu Asn Lys Asn Gln Glu Val Ile Leu Pro Asp
500 505 510

Glu Ile Pro Ile Val Asn Ile Leu Tyr Glu Tyr Leu Ile His Ile Asp
515 520 525

Lys Gln Ile Asp Arg Ser Ser Gly Leu Lys Leu Pro Phe Leu Arg Asn
530 535 540

Asp Phe Val Ser Ile Asn Glu Val Pro Thr Ala Ser Ser Leu Lys Trp
545 550 555 560

<210> 80
<211> 469
<212> PRT
<213> Kluyveromyces lactis

<400> 80

Met Ser Ala Leu Leu Arg Asp Ala Asp Leu Asn Asp Phe Ile Ser Pro
1 5 10 15

Gly Leu Ala Cys Val Lys Pro Ala Gln Pro Gln Lys Val Glu Lys Lys
20 25 30

Pro Ser Phe Glu Val Glu Val Gly Ile Glu Ser Ser Glu Pro Glu Lys
35 40 45

Val Ser Ile Ser Leu Gln Asp Cys Leu Ala Cys Ala Gly Cys Ile Thr
50 55 60

Ser Ser Glu Glu Ile Leu Leu Ser Lys Gln Ser His Lys Val Phe Leu
65 70 75 80

Glu Lys Trp Ser Glu Leu Glu Glu Leu Asp Glu Arg Ser Leu Ala Val
85 90 95

Ser Ile Ser Pro Gln Cys Arg Leu Ser Leu Ala Asp Tyr Tyr Ser Met
100 105 110

Cys Leu Ala Asp Leu Asp Arg Cys Phe Gln Asn Phe Met Lys Thr Lys
115 120 125

Phe Asn Ala Lys Tyr Val Val Gly Thr Gln Phe Gly Arg Ser Ile Ser
130 135 140

Ile Ser Arg Ile Asn Ala Thr Leu Lys Asp Arg Val Pro Glu Asn Glu
145 150 155 160

Gly Pro Leu Leu Cys Ser Val Cys Pro Gly Phe Val Leu Tyr Ala Glu
165 170 175

Lys Thr Lys Pro Glu Leu Ile Pro His Met Leu Asp Val Lys Ser Pro
180 185 190

Gln Gln Ile Thr Gly Asn Leu Leu Lys Gln Ala Asp Pro Thr Cys Tyr
195 200 205

His Leu Ser Ile Met Pro Cys Phe Asp Lys Lys Leu Glu Ala Ser Arg
210 215 220

Glu Glu Cys Glu Lys Glu Val Asp Cys Val Ile Thr Pro Lys Gln Phe
225 230 235 240

Val Ala Met Leu Gly Asp Leu Ser Ile Asp Phe Lys Ser Tyr Met Thr
245 250 255

Glu Tyr Asp Ser Ser Lys Glu Leu Cys Pro Ser Gly Trp Asp Tyr Lys
260 265 270

Leu His Trp Leu Ser Asn Glu Gly Ser Ser Ser Gly Gly Tyr Ala Tyr
275 280 285

Gln Tyr Leu Leu Ser Leu Gln Ser Ser Asn Pro Glu Ser Asp Ile Ile
290 295 300

Thr Ile Glu Gly Lys Asn Ser Asp Val Thr Glu Tyr Arg Leu Val Ser
305 310 315 320

Lys Ser Lys Gly Val Ile Ala Ser Ser Ser Glu Val Tyr Gly Phe Arg
325 330 335

Asn Ile Gln Asn Leu Val Arg Lys Leu Ser Gln Ser Ala Ser Val Lys
340 345 350

Lys Arg Gly Ile Lys Val Lys Arg Arg Gly Gln Ser Val Leu Lys Ser
355 360 365

Gly Glu Thr Ser Glu Lys Thr Thr Lys Val Leu Thr Ala Asp Pro Ala
370 375 380

Lys Thr Asp Phe Val Glu Val Met Ala Cys Pro Ser Gly Cys Ile Asn
385 390 395 400

Gly Gly Gly Leu Leu Asn Glu Glu Lys Asn Ala Asn Arg Arg Lys Gln
405 410 415

Leu Ala Gln Asp Leu Ser Leu Ala Tyr Thr Lys Val His Ser Val Asn
420 425 430

Ile Pro Asp Ile Val His Ala Tyr Asp Asp Lys Ser Asn Asp Phe Lys
435 440 445

Tyr Asn Leu Arg Val Ile Glu Pro Ser Thr Ser Ser Asp Val Val Ala
450 455 460

Val Gly Asn Thr Trp
465

<210> 81
<211> 365
<212> PRT
<213> Encephalitozoon cuniculi

<400> 81

Met Asp Ala Leu Ile Arg Pro Pro Met Ser Phe Phe Ala Asp Leu Pro
1 5 10 15

Lys Asp Asn Lys Lys Cys Ile Lys Ile Gly Ser Pro Leu Ala Leu Ser
20 25 30

Leu Ser Asp Cys Leu Ala Cys Ser Gly Cys Val Ser Ala Asp Glu Ala
35 40 45

Gly Ala Leu Ser Glu Asp Leu Ser Phe Val Leu Asp Leu Ser Pro Gln
50 55 60

Thr Ser Phe Val Leu Ser Pro Gln Ser Lys Ile Asn Ile Phe Asn Leu
65 70 75 80

Tyr Arg Glu Asp Gly Met Glu Tyr Arg Glu Phe Glu Ala Val Leu Ser
85 90 95

Ser Phe Leu Arg Ser Lys Phe Asn Ile His Arg Ile Val Asp Thr Ser
100 105 110

Tyr Leu Arg Ser Lys Ile Tyr Glu Glu Thr Tyr Arg Glu Tyr Met Ala
115 120 125

Thr Asn His Leu Ile Val Ser Ala Cys Pro Gly Val Val Thr Tyr Ile
130 135 140

Glu Arg Thr Ala Pro Tyr Leu Ile Gly Tyr Leu Ser Arg Val Lys Ser
145 150 155 160

Pro Gln Gln Met Ala Phe Ser Leu Val Lys Gly Ser Arg Thr Val Ser
165 170 175

Val Met Pro Cys Gln Asp Lys Lys Leu Glu Asn Gly Arg Asp Gly Val
180 185 190

Lys Phe Asp Phe Ile Leu Thr Thr Arg Gly Phe Cys Lys Ala Leu Asp
195 200 205

Ser Leu Gly Phe Arg Arg Pro Ala Arg Ala Ser Gly Lys Ser Leu Cys
210 215 220

Ser Met Glu Glu Ala Glu Thr Thr Gln Trp Asn Ile Gly Thr Ser Ser
225 230 235 240

Gly Gly Tyr Ala Glu Phe Ile Leu Gly Lys His Cys Val Val Glu Thr
245 250 255

Arg Glu Ile Arg Asn Gly Ile Lys Glu His Leu Leu Asp Asp Gly Arg
260 265 270

Thr Ile Ser Gln Ile Thr Gly Leu Glu Asn Ser Ile Asn Tyr Phe Lys
275 280 285

Ser Ser Lys Thr Lys Gly Pro Arg His Lys Met Thr Glu Ile Phe Leu
290 295 300

Cys Lys Asn Gly Cys Ile Gly Gly Pro Gly Gln Glu Arg Val Asn Asp
305 310 315 320

Val Glu Met Asp Ile Arg Glu Tyr Asp Arg Asn Gly Arg Glu Gln Pro
325 330 335

Arg Ile Phe Tyr Ser Ser Pro Gly Leu Glu Glu Lys Arg Val Phe Arg
340 345 350

Glu Val Lys Ala Lys Arg Val Asp Leu Arg Val Asp Trp
355 360 365

<210> 82
<211> 127
<212> PRT
<213> Tritrichomonas foetus

<220>
<221> misc_feature
<222> (85)..(85)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (124)..(124)
<223> Xaa can be any naturally occurring amino acid

<400> 82

Met Cys Ile Lys Ala Cys Asn Ser Val Ala Gly Gln Gly Val Leu Lys
1 5 10 15

Leu Val Lys Val Gly Asn Lys Lys Leu Val Ser Thr Lys Ser Gly Lys
20 25 30

Pro Leu Gln Glu Thr Asn Cys Ile Lys Cys Gly Gln Cys Thr Leu Val
35 40 45

Cys Gly Pro Gly Ala Leu Thr Gln Lys Asp Ala Ile Gln Thr Val Ser
50 55 60

Glu Val Leu Lys Asn Pro Gly Asp Lys Val Leu Val Cys Gln Thr Ala
65 70 75 80

Pro Ala Ile Arg Xaa Asn Leu Ala Asp Gly Leu Gly Met Pro Ala Gly
85 90 95

Ser Ile Ile Thr Gly Lys Met Val Thr Ala Leu Lys Met Leu Gly Phe
100 105 110

Lys Tyr Val Phe Asp Thr Asn Phe Gly Thr Asp Xaa Thr Ile Gly
115 120 125

<210> 83
<211> 449
<212> PRT
<213> Scenedesmus obliquus

<400> 83

Met Pro Glu Trp Gln Pro Gly Gly Arg Tyr Ala Val Ser Val Arg Pro
1 5 10 15

Pro Val Asn Arg Arg Ala Val Val Ala Ala Glu Arg Arg Arg Leu Val
20 25 30

Val Arg Ala Ala Gly Pro Thr Ala Glu Cys Asp Cys Pro Pro Ala Pro
35 40 45

Ala Pro Lys Ala Pro His Trp Gln Gln Thr Leu Asp Glu Leu Ala Lys
50 55 60

Pro Lys Glu Gln Arg Lys Val Met Ile Ala Gln Ile Ala Pro Ala Val
65 70 75 80

Arg Val Ala Ile Ala Glu Thr Met Gly Leu Asn Pro Gly Asp Val Thr
85 90 95

Val Gly Gln Met Val Thr Gly Leu Arg Met Leu Gly Phe Asp Tyr Val
100 105 110

Phe Asp Thr Leu Phe Gly Ala Asp Leu Thr Ile Met Glu Glu Gly Thr
115 120 125

Glu Leu Arg His Arg Leu Gln Asp His Leu Glu Gln His Pro Asn Lys
130 135 140

Glu Glu Pro Leu Pro Met Phe Thr Ser Cys Cys Pro Gly Trp Val Ala
145 150 155 160

Met Val Glu Lys Ser Asn Pro Glu Leu Ile Pro Tyr Leu Ser Ser Cys
165 170 175

Lys Ser Pro Gln Met Met Leu Gly Ala Val Ile Lys Asn Tyr Phe Ala
180 185 190

Ala Glu Ala Gly Ala Lys Pro Glu Asp Ile Cys Asn Val Ser Val Met
195 200 205

Pro Cys Val Arg Lys Gln Gly Glu Ala Asp Arg Glu Trp Phe Asn Thr
210 215 220

Thr Gly Ala Gly Gly Ala Asn Val Asp His Val Met Thr Thr Ala Glu
225 230 235 240

Leu Gly Lys Ile Phe Val Glu Arg Gly Ile Lys Leu Asn Asp Leu Gln
245 250 255

Glu Ser Pro Phe Asp Asn Pro Val Gly Glu Gly Ser Gly Gly Gly Val
260 265 270

Leu Phe Gly Thr Thr Gly Gly Val Met Glu Ala Ala Leu Arg Thr Val
275 280 285

Tyr Glu Val Val Thr Gln Lys Pro Leu Asp Arg Ile Val Phe Glu Asp
290 295 300

Val Arg Gly Leu Glu Gly Ile Lys Glu Ser Thr Leu His Leu Thr Pro
305 310 315 320

Gly Pro Thr Ser Pro Phe Lys Ala Phe Ala Gly Ala Asp Gly Thr Gly
325 330 335

Ile Thr Leu Asn Ile Ala Val Ala Asn Gly Leu Gly Asn Ala Lys Lys
340 345 350

Leu Ile Lys Gln Leu Ala Ala Gly Glu Ser Lys Tyr Asp Phe Ile Glu
355 360 365

Val Met Ala Cys Pro Gly Gly Cys Ile Gly Gly Gly Gly Gln Pro Arg
370 375 380

Ser Ala Asp Lys Gln Ile Leu Gln Lys Arg Gln Ala Ala Met Tyr Asp
385 390 395 400

Leu Asp Glu Arg Ala Val Ile Arg Arg Ser His Glu Asn Pro Leu Ile
405 410 415

Gly Ala Leu Tyr Glu Lys Phe Leu Gly Glu Pro Asn Gly His Lys Ala
420 425 430

His Glu Leu Leu His Thr His Tyr Val Ala Gly Gly Val Pro Asp Glu
435 440 445

Lys

<210> 84
<211> 477
<212> PRT
<213> Anopheles gambiae

<400> 84

Ser Arg Phe Ser Ser Ala Leu Gln Leu Thr Asp Leu Asp Asp Phe Ile
1 5 10 15

Thr Pro Ser Gln Glu Cys Ile Lys Pro Val Lys Ile Glu Thr Ser Lys
20 25 30

Ser Lys Thr Gly Ala Lys Ile Thr Ile Gln Glu Asp Gly Ser Tyr Val
35 40 45

Gln Glu Ser Ser Ser Gly Ile Gln Lys Leu Glu Lys Val Glu Ile Thr
50 55 60

Leu Ala Asp Cys Leu Ala Cys Ser Gly Cys Ile Thr Ser Ala Glu Gly
65 70 75 80

Val Leu Ile Ser Gln Gln Ser Gln Glu Glu Leu Leu Arg Val Met Asn
85 90 95

Ala Asn Asn Leu Ala Lys Leu Asn Asn Gln Arg Asp Glu Ile Lys Phe
100 105 110

Val Val Phe Thr Val Ser Gln Gln Pro Ile Leu Ser Leu Ala Arg Lys
115 120 125

Tyr Asn Leu Thr Pro Glu Asp Thr Phe Glu His Ile Ala Gly Tyr Phe
130 135 140

Lys Lys Leu Gly Ala Asp Met Val Val Asp Thr Lys Ile Ala Asp Asp
145 150 155 160

Leu Ala Leu Ile Glu Cys Arg Asn Glu Phe Ile Glu Arg Tyr Asn Thr
165 170 175

Asn Arg Lys Leu Leu Pro Met Leu Ala Ser Ser Cys Pro Gly Trp Val
180 185 190

Cys Tyr Ala Glu Lys Thr His Gly Asn Phe Ile Leu Pro Tyr Ile Ala
195 200 205

Thr Thr Arg Ser Pro Gln Gln Ile Met Gly Val Leu Val Lys Gln Tyr
210 215 220

Leu Ala Lys Gln Leu Gln Thr Thr Gly Asp Arg Ile Tyr His Val Thr
225 230 235 240

Val Met Pro Cys Tyr Asp Lys Lys Leu Glu Ala Ser Arg Glu Asp Phe
245 250 255

```
Phe Ser Glu Val Glu Asn Ser Arg Asp Val Asp Cys Val Ile Thr Ser
            260                 265                 270

Ile Glu Ile Glu Gln Met Leu Asn Ser Leu Asp Leu Pro Ser Leu Gln
        275                 280                 285

Leu Val Glu Arg Cys Ala Ile Asp Trp Pro Trp Pro Thr Val Arg Pro
    290                 295                 300

Ser Ala Phe Val Trp Gly His Glu Ser Ser Gly Ser Gly Gly Tyr Ala
305                 310                 315                 320

Glu Tyr Ile Phe Lys Tyr Ala Ala Arg Lys Leu Phe Asn Val Gln Leu
                325                 330                 335

Asp Thr Val Ala Phe Lys Pro Leu Arg Asn Asn Asp Met Arg Glu Ala
            340                 345                 350

Val Leu Glu Gln Asn Gly Gln Val Leu Met Arg Phe Ala Ile Ala Asn
        355                 360                 365

Gly Phe Arg Asn Ile Gln Asn Met Val Gln Lys Leu Lys Arg Gly Lys
    370                 375                 380

Ser Thr Tyr Asp Tyr Val Glu Ile Met Ala Cys Pro Ser Gly Cys Leu
385                 390                 395                 400

Asn Gly Gly Ala Gln Ile Arg Pro Glu Glu Gly Arg Ala Ala Arg Glu
                405                 410                 415

Leu Thr Ala Glu Leu Glu Cys Met Tyr Arg Ser Leu Pro Gln Ser Thr
            420                 425                 430

Pro Glu Asn Asp Cys Val Gln Thr Met Tyr Ala Thr Phe Phe Asp Ser
        435                 440                 445

Glu Gly Asp Leu Asn Lys Arg Gln Ser Leu Leu His Thr Ser Tyr His
    450                 455                 460

Gln Ile Glu Lys Ile Asn Ser Ala Leu Asn Ile Lys Trp
465                 470                 475

<210>  85
<211>  410
<212>  PRT
<213>  Shewanella oneidensis

<400>  85

Met Thr Thr Thr Thr Tyr Gln Pro Gly Glu Ile Gln Gly Leu Ile Lys
1               5                   10                  15

Ile Asn Ala Ser Lys Cys Lys Gly Cys Asp Ala Cys Lys Gln Phe Cys
            20                  25                  30

Pro Thr His Ala Ile Asn Gly Ala Ser Gly Ala Val His Ser Ile Asp
        35                  40                  45
```

Glu Asp Lys Cys Leu Ser Cys Gly Gln Cys Leu Ile Asn Cys Pro Phe
50 55 60

Ser Ala Ile Glu Glu Thr His Ser Ala Leu Glu Thr Val Ile Lys Lys
65 70 75 80

Leu Ala Asp Lys Asn Thr Thr Val Val Gly Ile Ile Ala Pro Ala Val
85 90 95

Arg Val Ala Ile Gly Glu Glu Phe Gly Leu Gly Thr Gly Glu Leu Val
100 105 110

Thr Gly Lys Leu Tyr Gly Ala Met Asn Gln Ala Gly Phe Lys Ile Phe
115 120 125

Asp Cys Asn Phe Ala Ala Asp Leu Thr Ile Met Glu Glu Gly Ser Glu
130 135 140

Phe Ile His Arg Leu His Ala Asn Val Lys Gly Glu Ala Asn Ala Gly
145 150 155 160

Pro Leu Pro Gln Phe Thr Ser Cys Cys Pro Gly Trp Val Arg Tyr Leu
165 170 175

Glu Thr Arg Tyr Pro Ala Leu Leu Pro Asn Leu Ser Thr Ala Lys Ser
180 185 190

Pro Gln Gln Met Ala Gly Thr Val Ala Lys Thr Tyr Gly Ala Lys Val
195 200 205

Tyr Gln Met Gln Pro Glu Asn Ile Phe Thr Val Ser Val Met Pro Cys
210 215 220

Thr Ser Lys Lys Leu Glu Ala Ser Arg Pro Glu Phe Asn Ser Ala Trp
225 230 235 240

Gln Tyr His Gln Glu His Gly Ala Asn Ser Pro Ser Tyr Gln Asp Ile
245 250 255

Asp Ala Val Leu Thr Thr Arg Glu Met Ala Gln Leu Leu Lys Leu Leu
260 265 270

Asp Ile Asp Leu Ala Asn Thr Ala Glu Tyr Gln Gly Asp Ser Leu Phe
275 280 285

Ser Glu Tyr Thr Gly Ala Gly Thr Ile Phe Gly Thr Thr Gly Gly Val
290 295 300

Met Glu Ala Ala Leu Arg Thr Ala His Lys Val Leu Thr Gly Thr Glu
305 310 315 320

Met Ala Lys Leu Glu Phe Glu Pro Val Arg Gly Leu Lys Gly Val Lys
325 330 335

Ser Ala Ser Val Ser Leu Phe Asp Thr Glu Leu Asn Gln Asp Val Thr
340 345 350

Val Asn Val Ala Val Val His Asp Met Gly Asn Asn Ile Glu Pro Val
355 360 365

Leu Arg Asp Val Met Ala Gly Thr Ser Pro Tyr His Phe Ile Glu Val
370 375 380

Met Asn Cys Ala Gly Gly Cys Val Asn Gly Gly Gly Gln Pro Ile Glu
385 390 395 400

Gly Lys Gly Ser Ser Trp Leu Gly Asn Ile
405 410

<210> 86
<211> 606
<212> PRT
<213> Clostridium thermocellum

<400> 86

Met Ala Phe Val Trp Arg Asn Val Arg Ser Arg Pro Phe Pro Lys Lys
1 5 10 15

Pro Asn Gly Arg Gly Cys Glu Lys Met Gln Met Val Asn Val Thr Ile
20 25 30

Asp Asn Cys Lys Ile Gln Val Pro Ala Asn Tyr Thr Val Leu Glu Ala
35 40 45

Ala Lys Gln Ala Asn Ile Asp Ile Pro Thr Leu Cys Phe Leu Lys Asp
50 55 60

Ile Asn Glu Val Gly Ala Cys Arg Met Cys Val Val Glu Val Lys Gly
65 70 75 80

Ala Arg Ser Leu Gln Ala Ala Cys Val Tyr Pro Val Ser Glu Gly Leu
85 90 95

Glu Val Tyr Thr Gln Thr Pro Ala Val Arg Glu Ala Arg Lys Val Thr
100 105 110

Leu Glu Leu Ile Leu Ser Asn His Glu Lys Lys Cys Leu Thr Cys Val
115 120 125

Arg Ser Glu Asn Cys Glu Leu Gln Arg Leu Ala Lys Asp Leu Asn Val
130 135 140

Lys Asp Ile Arg Phe Glu Gly Glu Met Ser Asn Leu Pro Ile Asp Asp
145 150 155 160

Leu Ser Pro Ser Val Val Arg Asp Pro Asn Lys Cys Val Leu Cys Arg
165 170 175

Arg Cys Val Ser Met Cys Lys Asn Val Gln Thr Val Gly Ala Ile Asp
180 185 190

Val Thr Glu Arg Gly Phe Arg Thr Thr Val Ser Thr Ala Phe Asn Lys
195 200 205

Pro Leu Ser Glu Val Pro Cys Val Asn Cys Gly Gln Cys Ile Asn Val
210 215 220

Cys Pro Val Gly Ala Leu Arg Glu Lys Asp Asp Ile Asp Lys Val Trp
225 230 235 240

Glu Ala Leu Ala Asn Pro Glu Leu His Val Val Val Gln Thr Ala Pro
245 250 255

Ala Val Arg Val Ala Leu Gly Glu Glu Phe Gly Met Pro Ile Gly Ser
260 265 270

Arg Val Thr Gly Lys Met Val Ala Ala Leu Ser Arg Leu Gly Phe Lys
275 280 285

Lys Val Phe Asp Thr Asp Thr Ala Ala Asp Leu Thr Ile Met Glu Glu
290 295 300

Gly Thr Glu Leu Ile Asn Arg Ile Lys Asn Gly Gly Lys Leu Pro Leu
305 310 315 320

Ile Thr Ser Cys Ser Pro Gly Trp Ile Lys Phe Cys Glu His Asn Tyr
325 330 335

Pro Glu Phe Leu Asp Asn Leu Ser Ser Cys Lys Ser Pro His Glu Met
340 345 350

Phe Gly Ala Val Leu Lys Ser Tyr Tyr Ala Gln Lys Asn Gly Ile Asp
355 360 365

Pro Ser Lys Val Phe Val Val Ser Ile Met Pro Cys Thr Ala Lys Lys
370 375 380

Phe Glu Ala Gln Arg Pro Glu Leu Ser Ser Thr Gly Tyr Pro Asp Val
385 390 395 400

Asp Val Val Leu Thr Thr Arg Glu Leu Ala Arg Met Ile Lys Glu Thr
405 410 415

Gly Ile Asp Phe Asn Ser Leu Pro Asp Lys Gln Phe Asp Asp Pro Met
420 425 430

Gly Glu Ala Ser Gly Ala Gly Val Ile Phe Gly Ala Thr Gly Gly Val
435 440 445

Met Glu Ala Ala Ile Arg Thr Val Gly Glu Leu Leu Ser Gly Lys Pro
450 455 460

Ala Asp Lys Ile Glu Tyr Thr Glu Val Arg Gly Leu Asp Gly Ile Lys
465 470 475 480

Glu Ala Ser Ile Glu Leu Asp Gly Phe Thr Leu Lys Ala Ala Val Ala
485 490 495

His Gly Leu Gly Asn Ala Arg Lys Leu Leu Asp Lys Ile Lys Ala Gly

505 510

Glu Ala Asp Tyr His Phe Ile Glu Ile Met Ala Cys Pro Gly Gly Cys
515 520 525

Ile Asn Gly Gly Gly Gln Pro Ile Gln Pro Ser Ser Val Arg Asn Trp
530 535 540

Lys Asp Ile Arg Cys Glu Arg Ala Lys Ala Ile Tyr Glu Glu Asp Glu
545 550 555 560

Ser Leu Pro Ile Arg Lys Ser His Glu Asn Pro Lys Ile Lys Met Leu
565 570 575

Tyr Glu Glu Phe Phe Gly Glu Pro Gly Ser His Lys Ala His Glu Leu
580 585 590

Leu His Thr His Tyr Glu Lys Arg Glu Asn Tyr Pro Val Lys
595 600 605

<210> 87
<211> 279
<212> PRT
<213> Desulfitobacterium hafniense

<400> 87

Met Thr Met Gly Gln Leu Arg Ala Ala Leu Lys His Leu Gly Phe Tyr
1 5 10 15

Gly Met Ile Glu Val Ala Leu Phe Ala Asp Val Leu Ser Leu Lys Glu
20 25 30

Ala Leu Glu Phe Asp Lys His Val Gln Thr Asp Lys Asp Phe Val Leu
35 40 45

Thr Ser Cys Cys Cys Pro Ile Trp Val Gly Met Val Lys Arg Val Tyr
50 55 60

Asp Thr Leu Val Pro His Ile Ser Pro Ser Val Ser Pro Met Val Ala
65 70 75 80

Cys Gly Arg Gly Ile Lys Arg Leu His Pro Asp Ala Lys Thr Val Phe
85 90 95

Ile Gly Pro Cys Ile Ala Lys Lys Ala Glu Ala Lys Glu Pro Asp Ile
100 105 110

Arg Asp Ala Val Asp Ala Val Leu Thr Phe His Glu Leu Lys Gln Ile
115 120 125

Phe Glu Ala Thr Asp Ile Glu Pro Ser Glu Met Glu Asp Ile Pro Ser
130 135 140

Glu His Ser Ser Thr Ser Gly Arg Ile Tyr Ala Arg Thr Gly Gly Val
145 150 155 160

Ser Lys Ser Ile Ser Asp Thr Leu Asn Arg Ile Arg Pro Asp Lys Pro

165 170 175

Val Lys Ile Lys Ser Ile Gln Ala Asn Gly Ile Lys Glu Cys Lys Ala
180 185 190

Leu Leu Asn Asp Ile Met Asn Asn Glu Ile Lys Ala Asn Phe Tyr Glu
195 200 205

Gly Met Gly Cys Pro Gly Gly Cys Val Gly Gly Pro Lys Ala Ile Val
210 215 220

Asp Val Asp Arg Gly Thr Glu Phe Val Asn Lys Tyr Gly Ala Glu Ala
225 230 235 240

Asp Ala Leu Thr Pro Ala Asp Asn Gln His Val Leu Glu Leu Leu Lys
245 250 255

Gln Leu Gly Ile Asp Ser Val Glu Glu Leu Leu Gly Gly Glu Ser Ala
260 265 270

Ala Ile Phe Gln Arg Asp Phe
275

<210> 88
<211> 505
<212> PRT
<213> C. reinhardtii

<400> 88

Met Ala Leu Gly Leu Arg Ala Glu Leu Arg Ala Gly Gln Ala Val Ala
1 5 10 15

Cys Ala Arg Arg Thr Asn Ala Pro Ala His Pro Ala Ala Val Val Pro
20 25 30

Val Leu Pro Ser Arg Gly Asp Lys Phe Phe Asn Leu Ser Gln Lys Val
35 40 45

Pro Ser Ser Gln Pro Ala Arg Gly Ser Thr Ile Arg Val Ala Ala Thr
50 55 60

Ala Thr Asp Ala Val Pro His Trp Lys Leu Ala Leu Glu Glu Leu Asp
65 70 75 80

Lys Pro Lys Asp Gly Gly Arg Lys Val Leu Ile Ala Gln Val Ala Pro
85 90 95

Ala Val Arg Val Ala Ile Ala Glu Ser Phe Gly Leu Ala Pro Gly Ala
100 105 110

Val Ser Pro Gly Lys Leu Ala Ala Gly Leu Arg Ala Leu Gly Phe Asp
115 120 125

Gln Val Phe Asp Thr Leu Phe Ala Ala Asp Leu Thr Ile Met Glu Glu
130 135 140

Gly Thr Glu Leu Leu His Arg Leu Lys Glu His Leu Glu Ala His Pro

His Ser Asp Glu Pro Leu Pro Met Phe Thr Ser Cys Cys Pro Gly Trp
165 170 175

Val Ala Met Met Glu Lys Ser Tyr Pro Glu Leu Ile Pro Phe Val Ser
180 185 190

Ser Cys Lys Ser Pro Gln Met Met Met Gly Ala Met Val Lys Thr Tyr
195 200 205

Leu Ser Glu Lys Gln Gly Ile Pro Ala Lys Asp Ile Val Met Val Ser
210 215 220

Val Met Pro Cys Val Arg Lys Gln Gly Glu Ala Asp Arg Glu Trp Phe
225 230 235 240

Cys Val Ser Glu Pro Gly Val Arg Asp Val Asp His Val Ile Thr Thr
245 250 255

Ala Glu Leu Gly Asn Ile Phe Lys Glu Arg Gly Ile Ile Leu Pro Glu
260 265 270

Leu Pro Asp Ser Asp Trp Asp Gln Pro Leu Gly Leu Gly Ser Gly Ala
275 280 285

Gly Val Leu Phe Gly Thr Thr Gly Gly Val Met Glu Ala Ala Val Arg
290 295 300

Thr Ala Tyr Glu Ile Val Thr Lys Glu Pro Leu Pro Arg Leu Asn Leu
305 310 315 320

Ser Glu Val Arg Gly Leu Asp Gly Ile Lys Glu Ala Ser Val Thr Leu
325 330 335

Val Pro Ala Pro Gly Ser Lys Phe Ala Glu Leu Val Ala Ala Arg Leu
340 345 350

Ala His Lys Val Glu Glu Ala Ala Ala Ala Glu Ala Ala Ala Ala Val
355 360 365

Glu Gly Ala Val Lys Pro Pro Ile Ala Tyr Asp Gly Gly Gln Gly Phe
370 375 380

Ser Thr Asp Asp Gly Lys Gly Gly Leu Lys Leu Arg Val Ala Val Ala
385 390 395 400

Asn Gly Leu Gly Asn Ala Lys Lys Leu Ile Gly Lys Met Val Ser Gly
405 410 415

Glu Ala Lys Tyr Asp Phe Val Glu Ile Met Ala Cys Pro Ala Gly Cys
420 425 430

Val Gly Gly Gly Gly Gln Pro Arg Ser Thr Asp Lys Gln Ile Thr Gln
435 440 445

Lys Arg Gln Ala Ala Leu Tyr Asp Leu Asp Glu Arg Asn Thr Leu Arg
450 455 460

Arg Ser His Glu Asn Glu Ala Val Asn Gln Leu Tyr Lys Glu Phe Leu
465 470 475 480

Gly Glu Pro Leu Ser His Arg Ala His Glu Leu Leu His Thr His Tyr
485 490 495

Val Pro Gly Gly Ala Glu Ala Asp Ala
500 505

<210> 89
<211> 505
<212> PRT
<213> C. reinhardtii

<400> 89

Met Ala Leu Gly Leu Arg Ala Glu Leu Arg Ala Gly Gln Ala Val Ala
1 5 10 15

Cys Ala Arg Arg Thr Asn Ala Pro Ala His Pro Ala Ala Val Val Pro
20 25 30

Val Leu Pro Ser Arg Gly Asp Lys Phe Phe Asn Leu Ser Gln Lys Val
35 40 45

Pro Ser Ser Gln Pro Ala Arg Gly Ser Thr Ile Arg Val Ala Ala Thr
50 55 60

Ala Thr Asp Ala Val Pro His Trp Lys Leu Ala Leu Glu Glu Leu Asp
65 70 75 80

Lys Pro Lys Asp Gly Gly Arg Lys Val Leu Ile Ala Gln Val Ala Pro
85 90 95

Ala Val Arg Val Ala Ile Ala Glu Ser Phe Gly Leu Ala Pro Gly Ala
100 105 110

Val Ser Pro Gly Lys Leu Ala Ala Gly Leu Arg Ala Leu Gly Phe Asp
115 120 125

Gln Val Phe Asp Thr Leu Phe Ala Ala Asp Leu Thr Ile Met Glu Glu
130 135 140

Gly Thr Glu Leu Leu His Arg Leu Lys Glu His Leu Glu Ala His Pro
145 150 155 160

His Ser Asp Glu Pro Leu Pro Met Phe Thr Ser Cys Cys Pro Gly Trp
165 170 175

Val Ala Met Met Glu Lys Ser Tyr Pro Glu Leu Ile Pro Phe Val Ser
180 185 190

Ser Cys Lys Ser Pro Gln Met Met Met Gly Ala Met Val Lys Thr Tyr
195 200 205

Leu Ser Glu Lys Gln Gly Ile Pro Ala Lys Asp Ile Val Met Val Ser
210 215 220

Val Met Pro Cys Val Arg Lys Gln Gly Val Ala Asp Arg Glu Trp Phe
225 230 235 240

Cys Val Ser Glu Pro Gly Val Arg Asp Val Asp His Val Ile Thr Thr
245 250 255

Ala Glu Leu Gly Asn Ile Phe Lys Glu Arg Gly Ile Ile Leu Pro Glu
260 265 270

Leu Pro Asp Ser Asp Trp Asp Gln Pro Leu Gly Leu Gly Ser Gly Ala
275 280 285

Gly Val Leu Phe Gly Thr Thr Gly Gly Val Met Glu Ala Ala Val Arg
290 295 300

Thr Ala Tyr Glu Ile Val Thr Lys Glu Pro Leu Pro Arg Leu Asn Leu
305 310 315 320

Ser Glu Val Arg Gly Leu Asp Gly Ile Lys Glu Ala Ser Val Thr Leu
325 330 335

Val Pro Ala Pro Gly Ser Lys Phe Ala Glu Leu Val Ala Ala Arg Leu
340 345 350

Ala His Lys Val Glu Glu Ala Ala Ala Ala Glu Ala Ala Ala Ala Val
355 360 365

Glu Gly Ala Val Lys Pro Pro Ile Ala Tyr Asp Gly Gly Gln Gly Phe
370 375 380

Ser Thr Asp Asp Gly Lys Gly Gly Leu Lys Leu Arg Val Ala Val Ala
385 390 395 400

Asn Gly Leu Gly Asn Ala Lys Lys Leu Ile Gly Lys Met Val Ser Gly
405 410 415

Glu Ala Lys Tyr Asp Phe Val Glu Ile Met Ala Cys Pro Ala Gly Cys
420 425 430

Val Gly Gly Gly Gly Gln Pro Arg Ser Thr Asp Lys Gln Ile Thr Gln
435 440 445

Lys Arg Gln Ala Ala Leu Tyr Asp Leu Asp Glu Arg Asn Thr Leu Arg
450 455 460

Arg Ser His Glu Asn Glu Ala Val Asn Gln Leu Tyr Lys Glu Phe Leu
465 470 475 480

Gly Glu Pro Leu Ser His Arg Ala His Glu Leu Leu His Thr His Tyr
485 490 495

Val Pro Gly Gly Ala Glu Ala Asp Ala
500 505

<210> 90
<211> 608
<212> PRT
<213> T. maritima

<400> 90

Met Arg Arg Phe Phe Lys Asn Asn Leu Arg Asn Leu Ser Gln Asn Gly
1 5 10 15

Glu Thr Asn Ser Val Arg Arg Cys Phe Ala Leu Ala Asp Val Thr Val
20 25 30

Val Ile Asn Gly Arg Thr Leu Thr Val Pro Asp Asn Leu Thr Val Ile
35 40 45

Glu Ala Cys Glu Lys Ala Gly Ile Glu Ile Pro Ala Leu Cys His His
50 55 60

Pro Arg Leu Gly Glu Ser Ile Gly Ala Cys Arg Val Cys Val Val Glu
65 70 75 80

Val Glu Gly Ala Arg Asn Leu Gln Pro Ala Cys Val Thr Lys Val Arg
85 90 95

Asp Gly Met Val Ile Lys Thr Ser Ser Asp Arg Val Lys Thr Ala Arg
100 105 110

Lys Phe Asn Leu Ala Leu Leu Leu Ser Glu His Pro Asn Asp Cys Met
115 120 125

Thr Cys Glu Ala Asn Gly Arg Cys Glu Phe Gln Asp Leu Ile Tyr Lys
130 135 140

Tyr Asp Val Glu Pro Ile Phe Gly Tyr Gly Thr Lys Glu Gly Leu Val
145 150 155 160

Asp Arg Ser Ser Pro Ala Ile Val Arg Asp Leu Ser Lys Cys Ile Lys
165 170 175

Cys Gln Arg Cys Val Arg Ala Cys Ser Glu Leu Gln Gly Met His Ile
180 185 190

Tyr Ser Met Val Glu Arg Gly His Arg Thr Tyr Pro Gly Thr Pro Phe
195 200 205

Asp Met Pro Val Tyr Glu Thr Asp Cys Ile Gly Cys Gly Gln Cys Ala
210 215 220

Ala Phe Cys Pro Thr Gly Ala Ile Val Glu Asn Ser Ala Val Lys Val
225 230 235 240

Val Leu Glu Glu Leu Glu Lys Lys Glu Lys Ile Leu Val Val Gln Thr
245 250 255

Ala Pro Ser Val Arg Val Ala Ile Gly Glu Glu Phe Gly Tyr Ala Pro
260 265 270

Gly Thr Ile Ser Thr Gly Gln Met Val Ala Ala Leu Arg Arg Leu Gly
275 280 285

Phe Asp Tyr Val Phe Asp Thr Asn Phe Gly Ala Asp Leu Thr Ile Met
290 295 300

Glu Glu Gly Ser Glu Phe Leu Glu Arg Leu Glu Lys Gly Asp Leu Glu
305 310 315 320

Asp Leu Pro Met Phe Thr Ser Cys Cys Pro Gly Trp Val Asn Leu Val
325 330 335

Glu Lys Val Tyr Pro Glu Leu Arg Thr Arg Leu Ser Ser Ala Lys Ser
340 345 350

Pro Gln Gly Met Leu Ser Ala Met Val Lys Thr Tyr Phe Ala Glu Lys
355 360 365

Leu Gly Val Lys Pro Glu Asp Ile Phe His Val Ser Ile Met Pro Cys
370 375 380

Thr Ala Lys Lys Asp Glu Ala Leu Arg Lys Gln Leu Met Val Asn Gly
385 390 395 400

Val Pro Ala Val Asp Val Val Leu Thr Thr Arg Glu Leu Gly Lys Leu
405 410 415

Ile Arg Met Lys Lys Ile Pro Phe Ala Asn Leu Pro Glu Glu Glu Tyr
420 425 430

Asp Ala Pro Leu Gly Ile Ser Thr Gly Ala Ala Ala Leu Phe Gly Val
435 440 445

Thr Gly Gly Val Met Glu Ala Ala Leu Arg Thr Ala Tyr Glu Leu Lys
450 455 460

Thr Gly Lys Ala Leu Pro Lys Ile Val Phe Glu Glu Val Arg Gly Leu
465 470 475 480

Lys Gly Val Arg Glu Ala Glu Ile Asp Leu Asp Gly Lys Lys Ile Arg
485 490 495

Ile Ala Val Val His Gly Thr Ala Asn Val Arg Asn Leu Val Glu Lys
500 505 510

Ile Leu Arg Arg Glu Val Lys Tyr His Phe Val Glu Val Met Ala Cys
515 520 525

Pro Gly Gly Cys Ile Gly Gly Gly Gly Gln Pro Tyr Ser Arg Asp Pro
530 535 540

Glu Ile Leu Arg Lys Arg Ala Glu Ala Ile Tyr Thr Ile Asp Glu Arg
545 550 555 560

Met Thr Leu Arg Lys Ser His Glu Asn Pro Ala Ile Lys Lys Leu Tyr
565 570 575

Glu Glu Tyr Leu Glu His Pro Leu Ser His Lys Ala His Glu Leu Leu
580 585 590

His Thr Tyr Tyr Glu Asp Arg Ser Arg Lys Lys Arg Leu Ala Val Lys
595 600 605

<210> 91
<211> 497
<212> PRT
<213> C. reinhardtii

<400> 91

Met Ser Ala Leu Val Leu Lys Pro Cys Ala Ala Val Ser Ile Arg Gly
1 5 10 15

Ser Ser Cys Arg Ala Arg Gln Val Ala Pro Arg Ala Pro Leu Ala Ala
20 25 30

Ser Thr Val Arg Val Ala Leu Ala Thr Leu Glu Ala Pro Ala Arg Arg
35 40 45

Leu Gly Asn Val Ala Cys Ala Ala Ala Ala Pro Ala Ala Glu Ala Pro
50 55 60

Leu Ser His Val Gln Gln Ala Leu Ala Glu Leu Ala Lys Pro Lys Asp
65 70 75 80

Asp Pro Thr Arg Lys His Val Cys Val Gln Val Ala Pro Ala Val Arg
85 90 95

Val Ala Ile Ala Glu Thr Leu Gly Leu Ala Pro Gly Ala Thr Thr Pro
100 105 110

Lys Gln Leu Ala Glu Gly Leu Arg Arg Leu Gly Phe Asp Glu Val Phe
115 120 125

Asp Thr Leu Phe Gly Ala Asp Leu Thr Ile Met Glu Glu Gly Ser Glu
130 135 140

Leu Leu His Arg Leu Thr Glu His Leu Glu Ala His Pro His Ser Asp
145 150 155 160

Glu Pro Leu Pro Met Phe Thr Ser Cys Cys Pro Gly Trp Ile Ala Met
165 170 175

Leu Glu Lys Ser Tyr Pro Asp Leu Ile Pro Tyr Val Ser Ser Cys Lys
180 185 190

Ser Pro Gln Met Met Leu Ala Ala Met Val Lys Ser Tyr Leu Ala Glu
195 200 205

Lys Lys Gly Ile Ala Pro Lys Asp Met Val Met Val Ser Ile Met Pro
210 215 220

Cys Thr Arg Lys Gln Ser Glu Ala Asp Arg Asp Trp Phe Cys Val Asp
225 230 235 240

Ala Asp Pro Thr Leu Arg Gln Leu Asp His Val Ile Thr Thr Val Glu
245 250 255

Leu Gly Asn Ile Phe Lys Glu Arg Gly Ile Asn Leu Ala Glu Leu Pro
260 265 270

Glu Gly Glu Trp Asp Asn Pro Met Gly Val Gly Ser Gly Ala Gly Val
275 280 285

Leu Phe Gly Thr Thr Gly Gly Val Met Glu Ala Ala Leu Arg Thr Ala
290 295 300

Tyr Glu Leu Phe Thr Gly Thr Pro Leu Pro Arg Leu Ser Leu Ser Glu
305 310 315 320

Val Arg Gly Met Asp Gly Ile Lys Glu Thr Asn Ile Thr Met Val Pro
325 330 335

Ala Pro Gly Ser Lys Phe Glu Glu Leu Leu Lys His Arg Ala Ala Ala
340 345 350

Arg Ala Glu Ala Ala Ala His Gly Thr Pro Gly Pro Leu Ala Trp Asp
355 360 365

Gly Gly Ala Gly Phe Thr Ser Glu Asp Gly Arg Gly Gly Ile Thr Leu
370 375 380

Arg Val Ala Val Ala Asn Gly Leu Gly Asn Ala Lys Lys Leu Ile Thr
385 390 395 400

Lys Met Gln Ala Gly Glu Ala Lys Tyr Asp Phe Val Glu Ile Met Ala
405 410 415

Cys Pro Ala Gly Cys Val Gly Gly Gly Gly Gln Pro Arg Ser Thr Asp
420 425 430

Lys Ala Ile Thr Gln Lys Arg Gln Ala Ala Leu Tyr Asn Leu Asp Glu
435 440 445

Lys Ser Thr Leu Arg Arg Ser His Glu Asn Pro Ser Ile Arg Glu Leu
450 455 460

Tyr Asp Thr Tyr Leu Gly Glu Pro Leu Gly His Lys Ala His Glu Leu
465 470 475 480

Leu His Thr His Tyr Val Ala Gly Gly Val Glu Glu Lys Asp Glu Lys
485 490 495

Lys

<210> 92
<211> 581
<212> PRT
<213> T. tencongensis

<400>192

Met Asp Lys Val Arg Val Thr Ile Asp Gly Ile Thr Val Glu Val Pro
1 5 10 15

Ser Tyr Tyr Thr Val Leu Glu Ala Ala Lys Glu Ala Gly Ile Asp Ile
20 25 30

Pro Thr Leu Cys Tyr Leu Lys Glu Ile Asn Gln Ile Gly Ala Cys Arg
35 40 45

Ile Cys Leu Val Glu Ile Glu Gly Val Arg Asn Leu Gln Thr Ser Cys
50 55 60

Thr Tyr Pro Val Phe Asp Gly Met Lys Val Tyr Thr Asn Thr Pro Lys
65 70 75 80

Ile Arg Glu Ala Arg Arg Leu Asn Leu Glu Leu Ile Leu Ser Asn His
85 90 95

Asp Arg Asn Cys Leu Thr Cys Val Arg Ser Thr Asn Cys Glu Leu Gln
100 105 110

Ala Leu Ala Lys Arg Leu Gly Val Glu Glu Ile Arg Phe Glu Gly Glu
115 120 125

Asn Ile Lys Tyr Pro Ile Asp Asp Ala Ser Pro Ala Val Val Arg Asp
130 135 140

Pro Asn Lys Cys Val Leu Cys Arg Arg Cys Val Ala Val Cys Ser Glu
145 150 155 160

Val Gln Asn Val Phe Ala Ile Gly Met Val Asn Arg Gly Phe Lys Thr
165 170 175

Met Val Ala Pro Ser Phe Gly Arg Ser Leu Lys Asp Ser Pro Cys Ile
180 185 190

Ser Cys Gly Gln Cys Ile Met Val Cys Pro Val Gly Ala Ile Tyr Glu
195 200 205

Lys Asp His Thr Lys Arg Val Tyr Glu Ala Leu Ala Asp Asp Lys Lys
210 215 220

Tyr Val Val Ala Gln Thr Ala Pro Ala Val Arg Val Ala Leu Gly Glu
225 230 235 240

Glu Phe Gly Met Pro Val Gly Thr Ile Val Thr Gly Lys Met Ala Ala
245 250 255

Ala Leu Arg Arg Met Gly Phe Asp Ala Val Phe Asp Thr Asn Phe Ala
260 265 270

Ala Asp Leu Thr Ile Met Glu Glu Gly Ser Glu Leu Leu Glu Arg Ile
275 280 285

Lys His Gly Gly Lys Leu Pro Met Ile Thr Ser Cys Ser Pro Gly Trp

290 295 300

Ile Ala Phe Cys Glu Lys Tyr Tyr Pro Glu Phe Ile Asp Asn Leu Ser
305 310 315 320

Thr Cys Lys Ser Pro His Met Met Met Gly Ala Leu Val Lys Ser Tyr
325 330 335

Tyr Ala Glu Lys Lys Gly Leu Asp Pro Lys Asp Ile Phe Val Val Ser
340 345 350

Ile Met Pro Cys Thr Ala Lys Lys Leu Glu Ile Glu Arg Glu Glu Met
355 360 365

Ile Arg Asn Gly Met Lys Asp Val Asp Ala Val Leu Thr Thr Arg Glu
370 375 380

Leu Ala Arg Met Ile Lys Glu Met Gly Ile Asp Phe Val Asn Leu Lys
385 390 395 400

Asp Glu Glu Phe Asp Glu Pro Leu Gly Met Ser Thr Gly Ala Gly Ala
405 410 415

Ile Phe Gly Ala Thr Gly Gly Val Met Glu Ala Ala Leu Arg Thr Val
420 425 430

Ala Glu Ile Val Glu Gly Arg Asp Ile Gly Lys Ile Asp Phe Glu Glu
435 440 445

Val Arg Gly Leu Glu Gly Val Arg Glu Ala Thr Ile Thr Ile Asp Gly
450 455 460

Met Asp Ile Lys Ile Ala Ile Ala Asn Gly Thr Gly Asn Ala Lys Lys
465 470 475 480

Leu Leu Asp Lys Val Lys Ala Gly Glu Val Glu Tyr His Phe Ile Glu
485 490 495

Val Met Gly Cys Pro Gly Gly Cys Ile Met Gly Gly Gly Gln Pro Ile
500 505 510

His Asn Pro Asn Glu Met Glu Glu Val Lys Lys Leu Arg Ala Lys Ala
515 520 525

Ile Tyr Glu Ile Asp Lys Asn Leu Pro Ile Arg Lys Ser His Glu Asn
530 535 540

Pro Ala Ile Lys Arg Leu Tyr Glu Glu Phe Leu Gly Tyr Pro Leu Ser
545 550 555 560

Glu Lys Ser His Glu Leu Leu His Thr His Tyr Ser Arg Lys Glu Leu
565 570 575

Tyr Pro Leu Val Lys
580

<210> 93
<211> 636
<212> PRT
<213> N. frontalis

<400> 93

Met Ser Met Leu Ser Ser Val Leu Asn Lys Ala Val Val Asn Pro Lys
1 5 10 15

Leu Thr Arg Ser Leu Ala Thr Ala Ala Ala Glu Lys Met Val Asn Ile
20 25 30

Ser Ile Asn Gly Arg Lys Phe Gln Val Lys Pro Lys Thr Thr Val Leu
35 40 45

Glu Ala Ala Lys Ala Asn Gly Tyr Tyr Ile Pro Thr Leu Cys Tyr His
50 55 60

Gln Glu Leu Pro Val Ala Gly Asn Cys Arg Leu Cys Leu Val Tyr Ala
65 70 75 80

Lys Gly Ser Trp Lys Pro Leu Thr Ala Cys Thr Thr Glu Val Trp Glu
85 90 95

Gly Met Glu Ile Glu Thr Asp Ser Pro Ala Val Ile Glu Thr Val Arg
100 105 110

Ser Ser Leu Ser Met Met Arg Glu Glu His Pro Asn Asp Cys Met Thr
115 120 125

Cys Gly Ser Asn Gly Asp Cys Glu Phe Gln Asp Leu Ile Tyr Arg Tyr
130 135 140

Gln Ile Asp Ala Lys His Pro Val Arg Ser Leu Leu Lys His Lys Ser
145 150 155 160

Lys Lys Thr Asn His Ser Ile Thr Glu Pro Cys Tyr Ser Pro Phe Asp
165 170 175

Asn Thr Thr Phe Ser Val Ala Arg Asp Met Asn Lys Cys Val Lys Cys
180 185 190

Gly Arg Cys Ile Arg Ala Cys His His Phe Gln Asn Ile Asn Ile Leu
195 200 205

Gly Phe Ile Asn Arg Ala Gly Tyr Glu Arg Val Gly Thr Pro Met Asp
210 215 220

Arg Pro Met Asn Phe Thr Lys Cys Val Glu Cys Gly Gln Cys Ser Gln
225 230 235 240

Val Cys Pro Val Gly Ala Ile Thr Ala Arg Thr Glu Val Val Asp Val
245 250 255

Leu Arg His Leu Asp Thr Lys Arg Lys Val Val Val Cys Ser Thr Ala
260 265 270

Pro Ala Ile Arg Val Ala Pro Ala Glu Glu Phe Ser Thr Glu Ala Asp
275 280 285

Phe Asp Phe Thr Gly Lys Met Val Ala Gly Leu Arg Lys Leu Gly Phe
290 295 300

Asp Tyr Ile Phe Asp Thr Asn Phe Ser Ala Asp Leu Thr Ile Met Glu
305 310 315 320

Glu Gly Thr Glu Leu Ile Asp Arg Leu Asn Asn Gly Gly Lys Phe Pro
325 330 335

Met Phe Thr Ser Cys Cys Pro Gly Trp Ile Asn Met Val Glu Lys Ser
340 345 350

Tyr Pro Glu Leu Ser Asp Asn Leu Ser Ser Cys Lys Ser Pro Gln Gln
355 360 365

Met Ile Gly Ala Val Ile Lys Ser Tyr Phe Ala Lys Lys Leu Gly Leu
370 375 380

Ser Thr Glu Asp Ile Ile His Val Ser Ile Met Pro Cys Thr Ala Lys
385 390 395 400

Lys Gly Glu Ala Arg Arg Pro Glu Phe Val Gln Lys Gly Lys Asp Gly
405 410 415

Lys Asp Tyr Pro Asp Ile Asp Tyr Val Ile Thr Thr Arg Glu Leu Leu
420 425 430

Thr Leu Leu Lys Leu Lys Lys Ile Asn Pro Ala Glu Leu Pro Asp Asp
435 440 445

Lys Phe Asp Ser Pro Leu Gly Ile Gly Ser Ser Ala Gly Asn Leu Phe
450 455 460

Gly Val Thr Gly Gly Val Met Glu Ala Ala Ile Arg Thr Ala Gln Val
465 470 475 480

Ile Thr Gly Val Glu Asn Pro Ile Pro Leu Gly Glu Leu Lys Ala Ile
485 490 495

Arg Gly Leu Asp Gly Ile Lys Ala Ala Asn Val Pro Leu Lys Thr Lys
500 505 510

Asp Gly Lys Glu Val Ser Val Arg Ala Ala Val Val Ser Gly Gly Ala
515 520 525

Asn Ile Gln Lys Phe Leu Glu Lys Ile Lys Asn Lys Glu Leu Glu Phe
530 535 540

Asp Phe Ile Glu Met Met Met Cys Pro Gly Gly Cys Ile Asn Gly Gly
545 550 555 560

Gly Gln Pro Lys Ser Ala Asp Pro Glu Ile Val Ala Lys Lys Met Gln
565 570 575

Arg Met Tyr Thr Met Asp Asp Gln Ala Lys Leu Arg Leu Cys His Glu
580 585 590

Asn Pro Glu Ile Ile Asp Val Tyr Lys Asn Phe Leu Gly Glu Pro Asn
595 600 605

Ser His Leu Ala His Glu Leu Leu His Thr His Tyr Asn Asp Arg Ser
610 615 620

Lys Thr Ile His Asp Met Gly His His Glu Lys Lys
625 630 635

<210> 94
<211> 579
<212> PRT
<213> C. thermocellum

<400> 94

Met Val Asn Val Thr Ile Asp Asn Cys Lys Ile Gln Val Pro Ala Asn
1 5 10 15

Tyr Thr Val Leu Glu Ala Ala Lys Gln Ala Asn Ile Asp Ile Pro Thr
20 25 30

Leu Cys Phe Leu Lys Asp Ile Asn Glu Val Gly Ala Cys Arg Met Cys
35 40 45

Val Val Glu Val Lys Gly Ala Arg Ser Leu Gln Ala Ala Cys Val Tyr
50 55 60

Pro Val Ser Glu Gly Leu Glu Val Tyr Thr Gln Thr Pro Ala Val Arg
65 70 75 80

Glu Ala Arg Lys Val Thr Leu Glu Leu Ile Leu Ser Asn His Glu Lys
85 90 95

Lys Cys Leu Thr Cys Val Arg Ser Glu Asn Cys Glu Leu Gln Arg Leu
100 105 110

Ala Lys Asp Leu Asn Val Lys Asp Ile Arg Phe Glu Gly Glu Met Ser
115 120 125

Asn Leu Pro Ile Asp Asp Leu Ser Pro Ser Val Val Arg Asp Pro Asn
130 135 140

Lys Cys Val Leu Cys Arg Arg Cys Val Ser Met Cys Lys Asn Val Gln
145 150 155 160

Thr Val Gly Ala Ile Asp Val Thr Glu Arg Gly Phe Arg Thr Thr Val
165 170 175

Ser Thr Ala Phe Asn Lys Pro Leu Ser Glu Val Pro Cys Val Asn Cys
180 185 190

Gly Gln Cys Ile Asn Val Cys Pro Val Gly Ala Leu Arg Glu Lys Asp
195 200 205

Asp Ile Asp Lys Val Trp Glu Ala Leu Ala Asn Pro Glu Leu His Val
210 215 220

Val Val Gln Thr Ala Pro Ala Val Arg Val Ala Leu Gly Glu Glu Phe
225 230 235 240

Gly Met Pro Ile Gly Ser Arg Val Thr Gly Lys Met Val Ala Ala Leu
245 250 255

Ser Arg Leu Gly Phe Lys Lys Val Phe Asp Thr Asp Thr Ala Ala Asp
260 265 270

Leu Thr Ile Met Glu Glu Gly Thr Glu Leu Ile Asn Arg Ile Lys Asn
275 280 285

Gly Gly Lys Leu Pro Leu Ile Thr Ser Cys Ser Pro Gly Trp Ile Lys
290 295 300

Phe Cys Glu His Asn Tyr Pro Glu Phe Leu Asp Asn Leu Ser Ser Cys
305 310 315 320

Lys Ser Pro His Glu Met Phe Gly Ala Val Leu Lys Ser Tyr Tyr Ala
325 330 335

Gln Lys Asn Gly Ile Asp Pro Ser Lys Val Phe Val Gly Ser Ile Met
340 345 350

Pro Cys Thr Ala Lys Lys Phe Glu Ala Gln Arg Pro Glu Leu Ser Ser
355 360 365

Thr Gly Tyr Pro Asp Val Asp Val Val Leu Thr Thr Arg Glu Leu Ala
370 375 380

Arg Met Ile Lys Glu Thr Gly Ile Asp Phe Asn Ser Leu Pro Asp Lys
385 390 395 400

Gln Phe Asp Asp Pro Met Gly Glu Ala Ser Gly Ala Gly Val Ile Phe
405 410 415

Gly Ala Thr Gly Gly Val Met Glu Ala Ala Ile Arg Thr Val Gly Glu
420 425 430

Leu Leu Ser Gly Lys Pro Ala Asp Lys Ile Glu Tyr Thr Glu Val Arg
435 440 445

Gly Leu Asp Gly Ile Lys Glu Ala Ser Ile Glu Leu Asp Gly Phe Thr
450 455 460

Leu Lys Ala Ala Val Ala His Gly Leu Gly Asn Ala Arg Lys Leu Leu
465 470 475 480

Asp Lys Ile Lys Ala Gly Glu Ala Asp Tyr His Phe Ile Glu Ile Met
485 490 495

Ala Cys Pro Gly Gly Cys Ile Asn Gly Gly Gly Gln Pro Ile Gln Pro
500 505 510

Ser Ser Val Arg Asn Trp Lys Asp Ile Arg Cys Glu Arg Ala Lys Ala
515 520 525

Ile Tyr Glu Glu Asp Glu Ser Leu Pro Ile Arg Lys Ser His Glu Asn
530 535 540

Pro Lys Ile Lys Met Leu Tyr Glu Glu Phe Phe Gly Glu Pro Gly Ser
545 550 555 560

His Lys Ala His Glu Leu Leu His Thr His Tyr Glu Lys Arg Glu Asn
565 570 575

Tyr Pro Val

<210> 95
<211> 588
<212> PRT
<213> B. thetaoimicron

<400> 95

Met Glu Glu Lys Gln Ile Thr Leu Gln Ile Asp Gly His Phe Ile Thr
1 5 10 15

Val Pro Glu Gly Ser Thr Ile Leu Glu Ala Ala Cys Lys Ile Gly Ile
20 25 30

Asn Ile Pro Thr Leu Cys His Ile Asp Leu Lys Gly Thr Cys Ile Lys
35 40 45

Asn Asn Pro Ala Ser Cys Arg Ile Cys Val Val Glu Val Ala Gly Arg
50 55 60

Arg Asn Leu Ala Pro Ala Cys Ala Thr Arg Cys Thr Glu Gly Met Val
65 70 75 80

Val Lys Thr Ser Thr Leu Arg Val Met Asn Ala Arg Lys Val Val Ala
85 90 95

Glu Leu Ile Leu Ser Asp His Pro Asn Asp Cys Leu Thr Cys Pro Lys
100 105 110

Cys Gly Asn Cys Glu Leu Gln Thr Leu Ala Leu Arg Phe Asn Ile Arg
115 120 125

Glu Met Pro Phe Asn Gly Gly Glu Leu Ser Pro Arg Lys Arg Glu Val
130 135 140

Thr Ser Ser Ile Val Arg Asn Met Asp Lys Cys Ile Phe Cys Arg Arg
145 150 155 160

Cys Glu Ser Val Cys Asn Asp Val Gln Thr Val Gly Ala Leu Gly Ala
165 170 175

Ile Arg Arg Gly Phe Asn Thr Thr Ile Ala Pro Ala Phe Asp Arg Met
180 185 190

Met Lys Asp Ser Glu Cys Thr Tyr Cys Gly Gln Cys Val Ala Val Cys
195 200 205

Pro Val Gly Ala Leu Thr Glu Arg Asp Tyr Thr Asn Arg Leu Leu Asp
210 215 220

Asp Leu Ala Asp Pro Asp Lys Ile Val Ile Val Gln Thr Ala Pro Ala
225 230 235 240

Val Arg Ala Ala Leu Gly Glu Glu Phe Gly Leu Pro Pro Gly Thr Leu
245 250 255

Val Thr Gly Lys Met Val Tyr Ala Leu Arg Glu Leu Gly Phe Asp Tyr
260 265 270

Val Phe Asp Thr Asp Phe Ala Ala Asp Leu Thr Ile Met Glu Glu Gly
275 280 285

Ser Glu Ile Leu Asn Arg Leu Thr Arg Tyr Leu Asp Gly Asp Lys Ser
290 295 300

Val Arg Leu Pro Ile Leu Thr Ser Cys Cys Pro Ala Trp Val Asn Phe
305 310 315 320

Phe Glu His His Phe Pro Asp Met Leu Asp Ile Pro Ser Thr Ala Arg
325 330 335

Ser Pro Gln Gln Met Phe Gly Ser Ile Ala Lys Ser Tyr Trp Ala Glu
340 345 350

Lys Met Gly Ile Pro Arg Glu Lys Leu Val Val Val Ser Ile Met Pro
355 360 365

Cys Leu Ala Lys Lys Tyr Glu Cys Asp Arg Asp Glu Phe Lys Val Asn
370 375 380

Gly Val Pro Asp Val Asp Tyr Ser Ile Ser Thr Arg Glu Leu Ala Arg
385 390 395 400

Leu Ile Lys Arg Ala Asn Ile Gly Phe Thr Leu Val Leu Asp Ser Pro
405 410 415

Phe Asp Asn Pro Met Gly Glu Ser Thr Gly Ala Gly Val Ile Phe Gly
420 425 430

Thr Thr Gly Gly Val Met Glu Ala Ala Leu Arg Ser Val Tyr Glu Ile
435 440 445

Tyr Thr Gly Gln Pro Leu Lys Asn Val Asn Phe Glu Gln Val Arg Gly
450 455 460

Leu Ser Gly Val Arg Arg Ala Thr Ile Asp Leu Asn Gly Phe Glu Leu
465 470 475 480

Lys Val Gly Ile Ala His Gly Leu Gly Asn Ala Arg His Leu Leu Glu

Asp Ile Arg Asn Gly His Asn Glu Tyr His Val Ile Glu Ile Met Ala
500 505 510

Cys Pro Gly Gly Cys Ile Gly Gly Gly Gly Gln Pro Leu His His Gly
515 520 525

Asn Ser Asp Val Leu Tyr Ala Arg Ala Asn Ala Leu Tyr Arg Glu Asp
530 535 540

Ala Asn Lys Pro Leu Arg Lys Ser His Asp Asn Pro Tyr Ile Gln Lys
545 550 555 560

Leu Tyr Glu Glu Tyr Leu Gly Lys Pro Leu Gly Glu Lys Ser Glu Met
565 570 575

Leu Leu His Thr His Tyr Phe Asn Lys Ser Ile Asp
580 585

<210> 96
<211> 585
<212> PRT
<213> D. fructosovorans

<400> 96

Met Ser Met Leu Thr Ile Thr Ile Asp Gly Lys Thr Thr Ser Val Pro
1 5 10 15

Glu Gly Ser Thr Ile Leu Asp Ala Ala Lys Thr Leu Asp Ile Asp Ile
20 25 30

Pro Thr Leu Cys Tyr Leu Asn Leu Glu Ala Leu Ser Ile Asn Asn Lys
35 40 45

Ala Ala Ser Cys Arg Val Cys Val Val Glu Val Glu Gly Arg Arg Asn
50 55 60

Leu Ala Pro Ser Cys Ala Thr Pro Val Thr Asp Asn Met Val Val Lys
65 70 75 80

Thr Asn Ser Leu Arg Val Leu Asn Ala Arg Arg Thr Val Leu Glu Leu
85 90 95

Leu Leu Ser Asp His Pro Lys Asp Cys Leu Val Cys Ala Lys Ser Gly
100 105 110

Glu Cys Glu Leu Gln Thr Leu Ala Glu Arg Phe Gly Ile Arg Glu Ser
115 120 125

Pro Tyr Asp Gly Gly Glu Met Ser His Tyr Arg Lys Asp Ile Ser Ala
130 135 140

Ser Ile Ile Arg Asp Met Asp Lys Cys Ile Met Cys Arg Arg Cys Glu
145 150 155 160

Thr Met Cys Asn Thr Val Gln Thr Cys Gly Val Leu Ser Gly Val Asn

165 170 175

Arg Gly Phe Thr Ala Val Val Ala Pro Ala Phe Glu Met Asn Leu Ala
180 185 190

Asp Thr Val Cys Thr Asn Cys Gly Gln Cys Val Ala Val Cys Pro Thr
195 200 205

Gly Ala Leu Val Glu His Glu Tyr Ile Trp Glu Val Val Glu Ala Leu
210 215 220

Ala Asn Pro Asp Lys Val Val Ile Val Gln Thr Ala Pro Ala Val Arg
225 230 235 240

Ala Ala Leu Gly Glu Asp Leu Gly Val Ala Pro Gly Thr Ser Val Thr
245 250 255

Gly Lys Met Ala Ala Ala Leu Arg Arg Leu Gly Phe Asp His Val Phe
260 265 270

Asp Thr Asp Phe Ala Ala Asp Leu Thr Ile Met Glu Glu Gly Ser Glu
275 280 285

Phe Leu Asp Arg Leu Gly Lys His Leu Ala Gly Asp Thr Asn Val Lys
290 295 300

Leu Pro Ile Leu Thr Ser Cys Cys Pro Gly Trp Val Lys Phe Phe Glu
305 310 315 320

His Gln Phe Pro Asp Met Leu Asp Val Pro Ser Thr Ala Lys Ser Pro
325 330 335

Gln Gln Met Phe Gly Ala Ile Ala Lys Thr Tyr Tyr Ala Asp Leu Leu
340 345 350

Gly Ile Pro Arg Glu Lys Leu Val Val Val Ser Val Met Pro Cys Leu
355 360 365

Ala Lys Lys Tyr Glu Cys Ala Arg Pro Glu Phe Ser Val Asn Gly Asn
370 375 380

Pro Asp Val Asp Ile Val Ile Thr Thr Arg Glu Leu Ala Lys Leu Val
385 390 395 400

Lys Arg Met Asn Ile Asp Phe Ala Gly Leu Pro Asp Glu Asp Phe Asp
405 410 415

Ala Pro Leu Gly Ala Ser Thr Gly Ala Ala Pro Ile Phe Gly Val Thr
420 425 430

Gly Gly Val Ile Glu Ala Ala Leu Arg Thr Ala Tyr Glu Leu Ala Thr
435 440 445

Gly Glu Thr Leu Lys Lys Val Asp Phe Glu Asp Val Arg Gly Met Asp
450 455 460

Gly Val Lys Lys Ala Lys Val Lys Val Gly Asp Asn Glu Leu Val Ile
465 470 475 480

Gly Val Ala His Gly Leu Gly Asn Ala Arg Glu Leu Leu Lys Pro Cys
485 490 495

Gly Ala Gly Glu Thr Phe His Ala Ile Glu Val Met Ala Cys Pro Gly
500 505 510

Gly Cys Ile Gly Gly Gly Gly Gln Pro Tyr His His Gly Asp Val Glu
515 520 525

Leu Leu Lys Lys Arg Thr Gln Val Leu Tyr Ala Glu Asp Ala Gly Lys
530 535 540

Pro Leu Arg Lys Ser His Glu Asn Pro Tyr Ile Ile Glu Leu Tyr Glu
545 550 555 560

Lys Phe Leu Gly Lys Pro Leu Ser Glu Arg Ser His Gln Leu Leu His
565 570 575

Thr His Tyr Phe Lys Arg Gln Arg Leu
580 585

<210> 97
<211> 606
<212> PRT
<213> D. vulgaris

<400> 97

Met Asn Ala Phe Ile Asn Gly Lys Glu Val Arg Cys Glu Pro Gly Arg
1 5 10 15

Thr Ile Leu Glu Ala Ala Arg Glu Asn Gly His Phe Ile Pro Thr Leu
20 25 30

Cys Glu Leu Ala Asp Ile Gly His Ala Pro Gly Thr Cys Arg Val Cys
35 40 45

Leu Val Glu Ile Trp Arg Asp Lys Glu Ala Gly Pro Gln Ile Val Thr
50 55 60

Ser Cys Thr Thr Pro Val Glu Glu Gly Met Arg Ile Phe Thr Arg Thr
65 70 75 80

Pro Glu Val Arg Arg Met Gln Arg Leu Gln Val Glu Leu Leu Leu Ala
85 90 95

Asp His Asp His Asp Cys Ala Ala Cys Ala Arg His Gly Asp Cys Glu
100 105 110

Leu Gln Asp Val Ala Gln Phe Val Gly Leu Thr Gly Thr Arg His His
115 120 125

Phe Pro Asp Tyr Ala Arg Ser Arg Thr Arg Asp Val Ser Ser Pro Ser
130 135 140

```
Val Val Arg Asp Met Gly Lys Cys Ile Arg Cys Leu Arg Cys Val Ala
145                 150                 155                 160

Val Cys Arg Asn Val Gln Gly Val Asp Ala Leu Val Val Thr Gly Asn
                165                 170                 175

Gly Ile Gly Thr Glu Ile Gly Leu Arg His Asn Arg Ser Gln Ser Ala
            180                 185                 190

Ser Asp Cys Val Gly Cys Gly Gln Cys Thr Leu Val Cys Pro Val Gly
        195                 200                 205

Ala Leu Ala Gly Arg Asp Asp Val Glu Arg Val Ile Asp Tyr Leu Tyr
    210                 215                 220

Asp Pro Glu Ile Val Thr Val Phe Gln Phe Ala Pro Ala Val Arg Val
225                 230                 235                 240

Gly Leu Gly Glu Glu Phe Gly Leu Pro Pro Gly Ser Ser Val Glu Gly
                245                 250                 255

Gln Val Pro Thr Ala Leu Arg Leu Leu Gly Ala Asp Val Val Leu Asp
            260                 265                 270

Thr Asn Phe Ala Ala Asp Leu Val Ile Met Glu Glu Gly Thr Glu Leu
        275                 280                 285

Leu Gln Arg Leu Arg Gly Gly Ala Lys Leu Pro Leu Phe Thr Ser Cys
    290                 295                 300

Cys Pro Gly Trp Val Asn Phe Ala Glu Lys His Leu Pro Asp Ile Leu
305                 310                 315                 320

Pro His Val Ser Thr Thr Arg Ser Pro Gln Gln Cys Leu Gly Ala Leu
                325                 330                 335

Ala Lys Thr Tyr Leu Ala Arg Thr Met Asn Val Ala Pro Glu Arg Met
            340                 345                 350

Arg Val Val Ser Leu Met Pro Cys Thr Ala Lys Lys Glu Glu Ala Ala
        355                 360                 365

Arg Pro Glu Phe Arg Arg Asp Gly Val Arg Asp Val Asp Ala Val Leu
    370                 375                 380

Thr Thr Arg Glu Phe Ala Arg Leu Leu Arg Arg Glu Gly Ile Asp Leu
385                 390                 395                 400

Ala Gly Leu Glu Pro Ser Pro Cys Asp Asp Pro Leu Met Gly Arg Ala
                405                 410                 415

Thr Gly Ala Ala Val Ile Phe Gly Thr Thr Gly Gly Val Met Glu Ala
            420                 425                 430

Ala Leu Arg Thr Val Tyr His Val Leu Asn Gly Lys Glu Leu Ala Pro
        435                 440                 445
```

Val Glu Leu His Ala Leu Arg Gly Tyr Glu Asn Val Arg Glu Ala Val
450 455 460

Val Pro Leu Gly Glu Gly Asn Gly Ser Val Lys Val Ala Val Val His
465 470 475 480

Gly Leu Lys Ala Ala Arg Gln Met Val Glu Ala Val Leu Ala Gly Lys
485 490 495

Ala Asp His Val Phe Val Glu Val Met Ala Cys Pro Gly Gly Cys Met
500 505 510

Asp Gly Gly Gly Gln Pro Arg Ser Lys Arg Ala Tyr Asn Pro Asn Ala
515 520 525

Gln Ala Arg Arg Ala Ala Leu Phe Ser Leu Asp Ala Glu Asn Ala Leu
530 535 540

Arg Gln Ser His Asn Asn Pro Leu Ile Gly Lys Val Tyr Glu Ser Phe
545 550 555 560

Leu Gly Glu Pro Cys Ser Asn Leu Ser His Arg Leu Leu His Thr Arg
565 570 575

Tyr Gly Asp Arg Lys Ser Glu Val Ala Tyr Thr Met Arg Asp Ile Trp
580 585 590

His Glu Met Thr Leu Gly Arg Arg Val Arg Gly Asp Ser Asp
595 600 605

<210> 98
<211> 589
<212> PRT
<213> T. vaginalis

<400> 98

Ala Ser Thr Gly Ile Asn Ser Thr Ala Asn Ile Leu Arg Asn Ile Thr
1 5 10 15

Val Thr Val Asn Gly Lys Pro Leu Glu Ala Lys Lys Gly Glu Thr Val
20 25 30

Leu Glu Leu Cys Asp Arg Asn Asn Ile Arg Ile Pro Arg Leu Cys Phe
35 40 45

His Pro Asn Leu Pro Pro Lys Ala Ser Cys Arg Val Cys Leu Val Glu
50 55 60

Cys Asp Gly Lys Trp Leu Ser Pro Ala Cys Val Thr Thr Val Trp Asp
65 70 75 80

Gly Leu Lys Ile Asp Thr Lys Ser Lys Asn Val Arg Asp Ser Val Glu
85 90 95

Asn Asn Leu Lys Glu Leu Leu Asp Cys His Asp Glu Thr Cys Ser Ala
100 105 110

Cys Ile Ala Asn His Arg Cys Gln Phe Arg Asp Met Asn Val Ala Tyr
115 120 125

Ser Val Lys Ala Glu Thr Lys Glu Ile Cys Ser Glu Glu Gly Ile Asp
130 135 140

Glu Ser Thr Asn Ala Ile Arg Leu Asp Thr Ser Lys Cys Val Leu Cys
145 150 155 160

Gly Arg Cys Ile Arg Ala Cys Glu Glu Val Ala Gly Thr Ser Ala Ile
165 170 175

Ile Phe Gly Asn Arg Ala Lys Lys Met Arg Ile Gln Pro Thr Phe Gly
180 185 190

Val Thr Leu Gln Glu Thr Ser Cys Ile Lys Cys Gly Gln Cys Thr Leu
195 200 205

Tyr Cys Pro Val Gly Ala Ile Thr Glu Lys Ser Gln Val Lys Glu Ala
210 215 220

Leu Asp Ile Leu Ala Asn Lys Gly Lys Lys Ile Thr Val Val Gln Val
225 230 235 240

Ala Pro Ala Val Arg Val Ala Leu Ser Glu Ala Phe Gly Tyr Lys Glu
245 250 255

Gly Thr Val Thr Thr Gly Lys Met Val Ser Ala Leu Lys Ala Leu Gly
260 265 270

Phe Asp Leu Val Tyr Asp Thr Asn Tyr Gly Ala Asp Leu Thr Ile Cys
275 280 285

Glu Glu Ala Gly Glu Leu Val Asn Arg Leu Arg Asp Pro Asn Ala Lys
290 295 300

Phe Pro Met Phe Thr Thr Cys Cys Pro Ala Trp Val Asn Tyr Val Glu
305 310 315 320

Gln Ser Ala Pro Asp Phe Ile Pro Asn Leu Ser Ser Cys Arg Ser Pro
325 330 335

Gln Gly Met Leu Ser Ala Leu Ile Lys Asn Tyr Leu Pro Lys Leu Leu
340 345 350

Asp Val Lys Gln Glu Asp Val Leu Asn Phe Ser Ile Met Pro Cys Thr
355 360 365

Ala Lys Lys Asp Glu Val Glu Arg Pro Glu Leu Arg Thr Lys Ser Gly
370 375 380

Leu Lys Glu Thr Asp Met Val Leu Thr Val Arg Glu Leu Val Glu Met
385 390 395 400

Ile Lys Leu Ser Asn Ile Asp Phe Asn Asn Leu Pro Asp Thr Gln Phe
405 410 415

Asp Asn Ile Phe Gly Phe Gly Ser Gly Ala Gly Gln Ile Phe Ala Ala
420 425 430

Thr Gly Gly Val Met Glu Ala Ala Ser Arg Thr Ala Phe Glu Val Tyr
435 440 445

Thr Gly Lys Lys Leu Thr Asn Val Asn Ile Tyr Pro Val Arg Gly Met
450 455 460

Asp Gly Leu Arg Ile Ala Glu Leu Asp Leu Asp Gly Thr Lys Leu Lys
465 470 475 480

Val Ala Val Cys His Gly Ile Ala Asn Thr Ala Lys Leu Leu Asp Arg
485 490 495

Leu Arg Glu Lys Asp Pro Glu Leu Met Asp Ile Lys Phe Ile Glu Ile
500 505 510

Met Ala Cys Pro Gly Gly Cys Val Cys Gly Gly Gly Thr Pro Gln Pro
515 520 525

Lys Asn Arg Val Ser Leu Asp Asn Arg Leu Ala Ala Ile Tyr Asn Ile
530 535 540

Asp Ala Lys Met Glu Cys Arg Lys Ser His Glu Asn Pro Leu Ile Lys
545 550 555 560

Gly Val Tyr Lys Glu Phe Leu Gly Lys Pro Asn Ser His Leu Ala His
565 570 575

Glu Leu Leu His Thr His Phe Lys His His Pro Lys Trp
580 585

<210> 99
<211> 1206
<212> PRT
<213> Nyctotherus ovalis

<400> 99

Met Ile Ser Arg Leu Ile Ala Lys Lys Ala Pro Leu Phe Leu Arg Thr
1 5 10 15

Phe Ala Thr Ser Glu Met Ile Ser Leu Lys Ile Asp Gly Lys Ile Ile
20 25 30

Ser Val Pro Lys Gly Ile Met Leu Ala Asp Ala Ile Lys Lys Ala Gly
35 40 45

Ala Asn Val Pro Thr Met Cys Tyr His Pro Asp Leu Pro Thr Ser Gly
50 55 60

Gly Ile Cys Arg Val Cys Leu Val Glu Ser Ala Lys Ser Pro Gly Tyr
65 70 75 80

Pro Ile Ile Ser Cys Arg Thr Pro Val Glu Glu Gly Met Glu Ile Val
85 90 95

```
Thr Gln Gly Ser Lys Met Lys Glu Tyr Arg Gln Ala Asn Leu Ala Leu
            100                 105                 110

Met Leu Ser Arg His Pro Asn Ala Cys Leu Ser Cys Thr Ser Asn Thr
        115                 120                 125

Asn Cys Lys Thr Gln Glu Leu Ser Ala Asn Met Asn Ile Gly Gln Cys
    130                 135                 140

Gly Phe Ala Asn Ala Thr Pro Pro Lys Asn Asp Asp Ser Tyr Asp Met
145                 150                 155                 160

Thr Thr Ala Ile Glu Arg Asp Asn Asp Lys Cys Ile Asn Cys Asp Ile
                165                 170                 175

Cys Val His Thr Cys Ser Leu Gln Gly Leu Asn Ala Leu Gly Phe Tyr
            180                 185                 190

Asn Glu Glu Gly His Ala Val Lys Ser Met Gly Thr Leu Asp Val Ser
        195                 200                 205

Glu Cys Ile Gln Cys Gly Gln Cys Ile Asn Arg Cys Pro Thr Gly Ala
    210                 215                 220

Ile Thr Glu Lys Ser Glu Ile Arg Pro Val Leu Asp Ala Ile Asn Ile
225                 230                 235                 240

Gln Gln Arg Leu Val Phe Gln Met Ala Pro Ser Ile Arg Val Ala Val
                245                 250                 255

Ala Glu Glu Phe Gly Ile Lys Pro Gly Glu Lys Ile Leu Lys Asn Glu
            260                 265                 270

Ile Ala Thr Ala Leu Arg Lys Leu Gly Ser Asn Val Phe Val Leu Asp
        275                 280                 285

Thr Asn Phe Ser Ala Asp Leu Thr Ile Ile Glu Glu Gly His Glu Leu
    290                 295                 300

Ile Glu Arg Leu Tyr Arg Asn Val Thr Gly Lys Lys Leu Leu Gly Gly
305                 310                 315                 320

Asp His Met Pro Ile Asp Leu Pro Met Leu Thr Ser Cys Cys Pro Gly
                325                 330                 335

Trp Ile Met Phe Ile Glu Lys Asn Tyr Pro Asp Leu Leu Asn Asn Leu
            340                 345                 350

Ser Thr Cys Lys Ser Pro Gln Gly Met Leu Gly Ala Leu Ile Lys Gly
        355                 360                 365

Tyr Trp Ala Lys Asn Ile Lys Lys Met Asp Pro Lys Asp Ile Val Ser
    370                 375                 380

Val Ser Ile Met Pro Cys Thr Ala Lys Lys Ala Glu Lys Glu Arg Pro
```

385 390 395 400

Gln Leu Arg Gly Asp Glu Gly Tyr Lys Asp Val Asp Tyr Ile Leu Thr
405 410 415

Thr Arg Glu Leu Ala Lys Met Leu Lys Gln Ser Asn Ile Asp Leu Ala
420 425 430

Lys Met Glu Pro Thr Pro Phe Asp Lys Val Met Ser Glu Gly Thr Gly
435 440 445

Ala Ala Val Ile Phe Gly Val Thr Gly Gly Val Met Glu Ala Ala Leu
450 455 460

Arg Thr Ala Asn Glu Val Ile Thr Gly Arg Glu Val Pro Phe Lys Asn
465 470 475 480

Leu Asn Ile Glu Ala Val Arg Gly Met Glu Gly Ile Arg Glu Ala Gly
485 490 495

Ile Lys Leu Glu Asn Val Leu Asp Lys Tyr Lys Ala Phe Glu Gly Val
500 505 510

Thr Val Lys Val Ala Ile Ala His Gly Pro Asn Asn Ala Arg Lys Val
515 520 525

Met Asp Ile Ile Lys Gln Ala Lys Glu Ser Gly Lys Pro Ala Pro Trp
530 535 540

His Phe Val Glu Val Met Ala Cys Pro Gly Gly Cys Ile Gly Gly Gly
545 550 555 560

Gly Gln Pro Lys Pro Thr Asn Leu Glu Ile Arg Gln Ala Arg Thr Gln
565 570 575

Leu Thr Phe Lys Glu Asp Met Asp Leu Pro Leu Arg Lys Ser His Asp
580 585 590

Asn Pro Glu Ile Lys Ala Ile Tyr Glu Asn Tyr Leu Lys Glu Pro Leu
595 600 605

Gly His Asn Ser His His Tyr Leu His Thr Thr Tyr Ser Ser Gln Lys
610 615 620

Val Arg Asp Met Asn Leu Tyr Asn Ala Asn Glu Ala Ala Gly Leu Asp
625 630 635 640

Glu Ile Leu Ala Lys Tyr Pro Lys Glu Lys Glu Tyr Leu Met Pro Ile
645 650 655

Ile Ile Glu Glu His Asp Lys Lys Gly Tyr Ile Ser Asp Pro Ser Ile
660 665 670

Val Lys Ile Ser Glu His Leu Gly Met Tyr Pro Ala Gln Ile Glu Ser
675 680 685

Ile Leu Ser Ser Tyr His Tyr Phe Pro Arg Glu His Thr Ile Ala Ile
690 695 700

Leu Met Ser Ile Cys Val His Cys His Asn Cys Met Met Lys Gly Gln
705 710 715 720

Gly Arg Leu Leu Lys Thr Ile Gln Glu Thr Tyr Asp Ile His Glu Thr
725 730 735

His Gly Gly Val Ala Lys Asp Gly Ser Phe Thr Leu His Thr Leu Asn
740 745 750

Trp Leu Gly Tyr Cys Val Asn Asp Ala Pro Ala Met Met Ile Lys Arg
755 760 765

Lys Gly Thr Asn Tyr Val Glu Thr Phe Thr Gly Leu Leu Gly Asp Asn
770 775 780

Ile Asp Gln Arg Leu Lys Ser Leu Lys Asn Leu Lys Lys Glu Leu Pro
785 790 795 800

Lys Trp Pro Lys Asn Asn Ile Arg Glu Met Lys Ser Gln Arg Asn Gly
805 810 815

Asn Ser Tyr Ser Cys Met Asn Thr Gln Ala Pro Ile Ala Glu Ala Thr
820 825 830

Lys Lys Ala Val Ser Met Gly Pro Glu Lys Val Ile Glu Glu Val Phe
835 840 845

Lys Ser Asn Leu Val Gly Arg Gly Gly Ala Gly Phe Arg Thr Gly Lys
850 855 860

Lys Trp Glu Ser Ala Tyr Lys Thr Pro Ala Ser Asp Lys Tyr Val Val
865 870 875 880

Cys Asn Ala Asp Glu Gly Leu Pro Ser Thr Tyr Lys Asp Trp Cys Leu
885 890 895

Leu Asn Asn Glu Ala Lys Arg Lys Glu Val Phe Thr Gly Met Gly Ile
900 905 910

Cys Ala Lys Thr Ile Gly Ala Lys Arg Cys Phe Met Tyr Leu Arg Tyr
915 920 925

Glu Tyr Arg Asn Leu Val Pro Ala Leu Glu Gln Ser Ile Lys Asp Val
930 935 940

Gln Ser Thr Cys Pro Glu Leu Ala Asp Leu Lys Tyr Glu Ile Arg Leu
945 950 955 960

Gly Gly Gly Pro Tyr Val Ala Gly Glu Glu Asn Ala Gln Phe Glu Ser
965 970 975

Ile Glu Gly Arg Ala Pro Leu Pro Arg Lys Asp Arg Pro Gly Asn Ile
980 985 990

Phe Pro Thr Met Glu Gly Leu Phe His Lys Pro Thr Val Ile Asn Asn
995 1000 1005

Val Glu Thr Phe Phe Ala Ile Pro His Ile Ile Gln Gln Gly Ser
1010 1015 1020

Gln Ser Phe Gly Glu Gly Lys Met Pro Lys Leu Leu Ser Val Thr
1025 1030 1035

Gly Asp Val Asp Glu Pro Ile Leu Ile Glu Thr Asn Leu Asn Asn
1040 1045 1050

Tyr Ser Leu Asn His Leu Leu Gln Glu Ile Ser Ala Lys Asp Ile
1055 1060 1065

Val Ala Ala Glu Ile Gly Gly Cys Thr Glu Pro Ile Ile Phe Gly
1070 1075 1080

Ser Lys Phe Asp Thr Leu Phe Gly Phe Gly Arg Gly Thr Leu Asn
1085 1090 1095

Ala Val Gly Ser Val Val Leu Phe Asn Ser Ser Cys Asp Leu Gly
1100 1105 1110

Lys Ile Tyr Glu Asn Lys Leu Lys Phe Met Ala Glu Glu Ser Cys
1115 1120 1125

Lys Gln Cys Val Pro Cys Arg Asp Gly Ser Tyr Ile Phe His Arg
1130 1135 1140

Ala Phe Lys Glu Leu Arg Asp Thr Gly Lys Ser Ser Tyr Asn Met
1145 1150 1155

Arg Ala Leu Ala Val Ala Ser Glu Ser Ala Ala Arg Ser Ser Ile
1160 1165 1170

Cys Ala His Gly Lys Ala Leu Glu Ser Leu Phe Lys Ser Ala Cys
1175 1180 1185

Asp Phe Met Asn Lys Thr Lys Pro Ile Tyr Gln Pro His Ser Thr
1190 1195 1200

Tyr His Gln
1205

<210> 100
<211> 468
<212> PRT
<213> T vaginalis

<400> 100

Met Leu Ala Ser Ser Ala Thr Ala Met Lys Gly Phe Ala Asn Ser Leu
1 5 10 15

Arg Met Lys Asp Tyr Ser Ser Thr Gly Ile Asn Phe Asp Met Thr Lys
20 25 30

Cys Ile Asn Cys Gln Ser Cys Val Arg Ala Cys Thr Asn Ile Ala Gly
35 40 45

Gln Asn Val Leu Lys Ser Leu Thr Val Asn Gly Lys Ser Val Val Gln
50 55 60

Thr Val Thr Gly Lys Pro Leu Ala Glu Thr Asn Cys Ile Ser Cys Gly
65 70 75 80

Gln Cys Thr Leu Gly Cys Pro Lys Phe Thr Ile Phe Glu Ala Asp Ala
85 90 95

Ile Asn Pro Val Lys Glu Val Leu Thr Lys Lys Asn Gly Arg Ile Ala
100 105 110

Val Cys Gln Ile Ala Pro Ala Ile Arg Ile Asn Met Ala Glu Ala Leu
115 120 125

Gly Val Pro Ala Gly Thr Ile Ser Leu Gly Lys Val Val Thr Ala Leu
130 135 140

Lys Arg Leu Gly Phe Asp Tyr Val Phe Asp Thr Asn Phe Ala Ala Asp
145 150 155 160

Met Thr Ile Val Glu Glu Ala Thr Glu Leu Val Gln Arg Leu Ser Asp
165 170 175

Lys Asn Ala Val Leu Pro Met Phe Thr Ser Cys Cys Pro Ala Trp Val
180 185 190

Asn Tyr Val Glu Lys Ser Asp Pro Ser Leu Ile Pro Tyr Leu Ser Ser
195 200 205

Cys Arg Ser Pro Met Ser Met Leu Ser Ser Val Ile Lys Asn Val Phe
210 215 220

Pro Lys Lys Ile Gly Thr Thr Ala Asp Lys Ile Tyr Asn Val Ala Ile
225 230 235 240

Met Pro Cys Thr Arg Lys Lys Asp Glu Ile Gln Arg Ser Gln Phe Thr
245 250 255

Met Lys Asp Gly Lys Gln Glu Thr Gly Ala Val Leu Thr Ser Arg Glu
260 265 270

Leu Ala Lys Met Ile Lys Glu Ala Lys Ile Asn Phe Lys Glu Leu Pro
275 280 285

Asp Thr Pro Cys Asp Asn Phe Tyr Ser Glu Ala Ser Gly Gly Gly Ala
290 295 300

Ile Phe Cys Ala Thr Gly Gly Val Met Glu Ala Ala Val Arg Ser Ala
305 310 315 320

Tyr Lys Phe Leu Thr Lys Lys Glu Leu Ala Pro Ile Asp Leu Gln Asp
325 330 335

Val Arg Gly Val Ala Ser Gly Val Lys Leu Ala Glu Val Asp Ile Ala
340 345 350

Gly Thr Lys Val Lys Val Ala Val Ala His Gly Ile Lys Asn Ala Met
355 360 365

Thr Leu Ile Lys Lys Ile Lys Ser Gly Glu Glu Gln Phe Lys Asp Val
370 375 380

Lys Phe Val Glu Val Met Ala Cys Pro Gly Gly Cys Val Val Gly Gly
385 390 395 400

Gly Ser Pro Lys Ala Lys Thr Lys Lys Ala Val Gln Ala Arg Leu Asn
405 410 415

Ala Thr Tyr Ser Ile Asp Lys Ser Ser Lys His Arg Thr Ser Gln Asp
420 425 430

Asn Pro Gln Leu Leu Gln Leu Tyr Lys Glu Ser Phe Glu Gly Lys Phe
435 440 445

Gly Gly His Val Ala His His Leu Leu His Thr His Tyr Lys Asn Arg
450 455 460

Lys Val Asn Pro
465

<210> 101
<211> 582
<212> PRT
<213> C. acetobutylicum

<400> 101

Met Lys Thr Ile Ile Leu Asn Gly Asn Glu Val His Thr Asp Lys Asp
1 5 10 15

Ile Thr Ile Leu Glu Leu Ala Arg Glu Asn Asn Val Asp Ile Pro Thr
20 25 30

Leu Cys Phe Leu Lys Asp Cys Gly Asn Phe Gly Lys Cys Gly Val Cys
35 40 45

Met Val Glu Val Glu Gly Lys Gly Phe Arg Ala Ala Cys Val Ala Lys
50 55 60

Val Glu Asp Gly Met Val Ile Asn Thr Glu Ser Asp Glu Val Lys Glu
65 70 75 80

Arg Ile Lys Lys Arg Val Ser Met Leu Leu Asp Lys His Glu Phe Lys
85 90 95

Cys Gly Gln Cys Ser Arg Arg Glu Asn Cys Glu Phe Leu Lys Leu Val
100 105 110

Ile Lys Thr Lys Ala Lys Ala Ser Lys Pro Phe Leu Pro Glu Asp Lys
115 120 125

Asp Ala Leu Val Asp Asn Arg Ser Lys Ala Ile Val Ile Asp Arg Ser
130 135 140

Lys Cys Val Leu Cys Gly Arg Cys Val Ala Ala Cys Lys Gln His Thr
145 150 155 160

Ser Thr Cys Ser Ile Gln Phe Ile Lys Lys Asp Gly Gln Arg Ala Val
165 170 175

Gly Thr Val Asp Asp Val Cys Leu Asp Asp Ser Thr Cys Leu Leu Cys
180 185 190

Gly Gln Cys Val Ile Ala Cys Pro Val Ala Ala Leu Lys Glu Lys Ser
195 200 205

His Ile Glu Lys Val Gln Glu Ala Leu Asn Asp Pro Lys Lys His Val
210 215 220

Ile Val Ala Met Ala Pro Ser Val Arg Thr Ala Met Gly Glu Leu Phe
225 230 235 240

Lys Met Gly Tyr Gly Lys Asp Val Thr Gly Lys Leu Tyr Thr Ala Leu
245 250 255

Arg Met Leu Gly Phe Asp Lys Val Phe Asp Ile Asn Phe Gly Ala Asp
260 265 270

Met Thr Ile Met Glu Glu Ala Thr Glu Leu Leu Gly Arg Val Lys Asn
275 280 285

Asn Gly Pro Phe Pro Met Phe Thr Ser Cys Cys Pro Ala Trp Val Arg
290 295 300

Leu Ala Gln Asn Tyr His Pro Glu Leu Leu Asp Asn Leu Ser Ser Ala
305 310 315 320

Lys Ser Pro Gln Gln Ile Phe Gly Thr Ala Ser Lys Thr Tyr Tyr Pro
325 330 335

Ser Ile Ser Gly Ile Ala Pro Glu Asp Val Tyr Thr Val Thr Ile Met
340 345 350

Pro Cys Asn Asp Lys Lys Tyr Glu Ala Asp Ile Pro Phe Met Glu Thr
355 360 365

Asn Ser Leu Arg Asp Ile Asp Ala Ser Leu Thr Thr Arg Glu Leu Ala
370 375 380

Lys Met Ile Lys Asp Ala Lys Ile Lys Phe Ala Asp Leu Glu Asp Gly
385 390 395 400

Glu Val Asp Pro Ala Met Gly Thr Tyr Ser Gly Ala Gly Ala Ile Phe
405 410 415

Gly Ala Thr Gly Gly Val Met Glu Ala Ala Ile Arg Ser Ala Lys Asp

420 425 430

Phe Ala Glu Asn Lys Glu Leu Glu Asn Val Asp Tyr Thr Glu Val Arg
435 440 445

Gly Phe Lys Gly Ile Lys Glu Ala Glu Val Glu Ile Ala Gly Asn Lys
450 455 460

Leu Asn Val Ala Val Ile Asn Gly Ala Ser Asn Phe Phe Glu Phe Met
465 470 475 480

Lys Ser Gly Lys Met Asn Glu Lys Gln Tyr His Phe Ile Glu Val Met
485 490 495

Ala Cys Pro Gly Gly Cys Ile Asn Gly Gly Gly Gln Pro His Val Asn
500 505 510

Ala Leu Asp Arg Glu Asn Val Asp Tyr Arg Lys Leu Arg Ala Ser Val
515 520 525

Leu Tyr Asn Gln Asp Lys Asn Val Leu Ser Lys Arg Lys Ser His Asp
530 535 540

Asn Pro Ala Ile Ile Lys Met Tyr Asp Ser Tyr Phe Gly Lys Pro Gly
545 550 555 560

Glu Gly Leu Ala His Lys Leu Leu His Val Lys Tyr Thr Lys Asp Lys
565 570 575

Asn Val Ser Lys His Glu
580

<210> 102
<211> 574
<212> PRT
<213> Clostridium pasteurianum

<400> 102

Met Lys Thr Ile Ile Ile Asn Gly Val Gln Phe Asn Thr Asp Glu Asp
1 5 10 15

Thr Thr Ile Leu Lys Phe Ala Arg Asp Asn Asn Ile Asp Ile Ser Ala
20 25 30

Leu Cys Phe Leu Asn Asn Cys Asn Asn Asp Ile Asn Lys Cys Glu Ile
35 40 45

Cys Thr Val Glu Val Glu Gly Thr Gly Leu Val Thr Ala Cys Asp Thr
50 55 60

Leu Ile Glu Asp Gly Met Ile Ile Asn Thr Asn Ser Asp Ala Val Asn
65 70 75 80

Glu Lys Ile Lys Ser Arg Ile Ser Gln Leu Leu Asp Ile His Glu Phe
85 90 95

Lys Cys Gly Pro Cys Asn Arg Arg Glu Asn Cys Glu Phe Leu Lys Leu

100 105 110

Val Ile Lys Tyr Lys Ala Arg Ala Ser Lys Pro Phe Leu Pro Lys Asp
115 120 125

Lys Thr Glu Tyr Val Asp Glu Arg Ser Lys Ser Leu Thr Val Asp Arg
130 135 140

Thr Lys Cys Leu Leu Cys Gly Arg Cys Val Asn Ala Cys Gly Lys Asn
145 150 155 160

Thr Glu Thr Tyr Ala Met Lys Phe Leu Asn Lys Asn Gly Lys Thr Ile
165 170 175

Ile Gly Ala Glu Asp Glu Lys Cys Phe Asp Asp Thr Asn Cys Leu Leu
180 185 190

Cys Gly Gln Cys Ile Ile Ala Cys Pro Val Ala Ala Leu Ser Glu Lys
195 200 205

Ser His Met Asp Arg Val Lys Asn Ala Leu Asn Ala Pro Glu Lys His
210 215 220

Val Ile Val Ala Met Ala Pro Ser Val Arg Ala Ser Ile Gly Glu Leu
225 230 235 240

Phe Asn Met Gly Phe Gly Val Asp Val Thr Gly Lys Ile Tyr Thr Ala
245 250 255

Leu Arg Gln Leu Gly Phe Asp Lys Ile Phe Asp Ile Asn Phe Gly Ala
260 265 270

Asp Met Thr Ile Met Glu Glu Ala Thr Glu Leu Val Gln Arg Ile Glu
275 280 285

Asn Asn Gly Pro Phe Pro Met Phe Thr Ser Cys Cys Pro Gly Trp Val
290 295 300

Arg Gln Ala Glu Asn Tyr Tyr Pro Glu Leu Leu Asn Asn Leu Ser Ser
305 310 315 320

Ala Lys Ser Pro Gln Gln Ile Phe Gly Thr Ala Ser Lys Thr Tyr Tyr
325 330 335

Pro Ser Ile Ser Gly Leu Asp Pro Lys Asn Val Phe Thr Val Thr Val
340 345 350

Met Pro Cys Thr Ser Lys Lys Phe Glu Ala Asp Arg Pro Gln Met Glu
355 360 365

Lys Asp Gly Leu Arg Asp Ile Asp Ala Val Ile Thr Thr Arg Glu Leu
370 375 380

Ala Lys Met Ile Lys Asp Ala Lys Ile Pro Phe Ala Lys Leu Glu Asp
385 390 395 400

Ser Glu Ala Asp Pro Ala Met Gly Glu Tyr Ser Gly Ala Gly Ala Ile
405 410 415

Phe Gly Ala Thr Gly Gly Val Met Glu Ala Ala Leu Arg Ser Ala Lys
420 425 430

Asp Phe Ala Glu Asn Ala Glu Leu Glu Asp Ile Glu Tyr Lys Gln Val
435 440 445

Arg Gly Leu Asn Gly Ile Lys Glu Ala Glu Val Glu Ile Asn Asn Asn
450 455 460

Lys Tyr Asn Val Ala Val Ile Asn Gly Ala Ser Asn Leu Phe Lys Phe
465 470 475 480

Met Lys Ser Gly Met Ile Asn Glu Lys Gln Tyr His Phe Ile Glu Val
485 490 495

Met Ala Cys His Gly Gly Cys Val Asn Gly Gly Gly Gln Pro His Val
500 505 510

Asn Pro Lys Asp Leu Glu Lys Val Asp Ile Lys Lys Val Arg Ala Ser
515 520 525

Val Leu Tyr Asn Gln Asp Glu His Leu Ser Lys Arg Lys Ser His Glu
530 535 540

Asn Thr Ala Leu Val Lys Met Tyr Gln Asn Tyr Phe Gly Lys Pro Gly
545 550 555 560

Glu Gly Arg Ala His Glu Ile Leu His Phe Lys Tyr Lys Lys
565 570

<210> 103
<211> 421
<212> PRT
<213> Desulfovibrio vulgaris

<400> 103

Met Ser Arg Thr Val Met Glu Arg Ile Glu Tyr Glu Met His Thr Pro
1 5 10 15

Asp Pro Lys Ala Asp Pro Asp Lys Leu His Phe Val Gln Ile Asp Glu
20 25 30

Ala Lys Cys Ile Gly Cys Asp Thr Cys Ser Gln Tyr Cys Pro Thr Ala
35 40 45

Ala Ile Phe Gly Glu Met Gly Glu Pro His Ser Ile Pro His Ile Glu
50 55 60

Ala Cys Ile Asn Cys Gly Gln Cys Leu Thr His Cys Pro Glu Asn Ala
65 70 75 80

Ile Tyr Glu Ala Gln Ser Trp Val Pro Glu Val Glu Lys Lys Leu Lys
85 90 95

Asp Gly Lys Val Lys Cys Ile Ala Met Pro Ala Pro Ala Val Arg Tyr
100 105 110

Ala Leu Gly Asp Ala Phe Gly Met Pro Val Gly Ser Val Thr Thr Gly
115 120 125

Lys Met Leu Ala Ala Leu Gln Lys Leu Gly Phe Ala His Cys Trp Asp
130 135 140

Thr Glu Phe Thr Ala Asp Val Thr Ile Trp Glu Glu Gly Ser Glu Phe
145 150 155 160

Val Glu Arg Leu Thr Lys Lys Ser Asp Met Pro Leu Pro Gln Phe Thr
165 170 175

Ser Cys Cys Pro Gly Trp Gln Lys Tyr Ala Glu Thr Tyr Tyr Pro Glu
180 185 190

Leu Leu Pro His Phe Ser Thr Cys Lys Ser Pro Ile Gly Met Asn Gly
195 200 205

Ala Leu Ala Lys Thr Tyr Gly Ala Glu Arg Met Lys Tyr Asp Pro Lys
210 215 220

Gln Val Tyr Thr Val Ser Ile Met Pro Cys Ile Ala Lys Lys Tyr Glu
225 230 235 240

Gly Leu Arg Pro Glu Leu Lys Ser Ser Gly Met Arg Asp Ile Asp Ala
245 250 255

Thr Leu Thr Thr Arg Glu Leu Ala Tyr Met Ile Lys Lys Ala Gly Ile
260 265 270

Asp Phe Ala Lys Leu Pro Asp Gly Lys Arg Asp Ser Leu Met Gly Glu
275 280 285

Ser Thr Gly Gly Ala Thr Ile Phe Gly Val Thr Gly Gly Val Met Glu
290 295 300

Ala Ala Leu Arg Phe Ala Tyr Glu Ala Val Thr Gly Lys Lys Pro Asp
305 310 315 320

Ser Trp Asp Phe Lys Ala Val Arg Gly Leu Asp Gly Ile Lys Glu Ala
325 330 335

Thr Val Asn Val Gly Gly Thr Asp Val Lys Val Ala Val Val His Gly
340 345 350

Ala Lys Arg Phe Lys Gln Val Cys Asp Asp Val Lys Ala Gly Lys Ser
355 360 365

Pro Tyr His Phe Ile Glu Tyr Met Ala Cys Pro Gly Gly Cys Val Cys
370 375 380

Gly Gly Gly Gln Pro Val Met Pro Gly Val Leu Glu Ala Met Asp Arg
385 390 395 400

Thr Thr Thr Arg Leu Tyr Ala Gly Leu Lys Lys Arg Leu Ala Met Ala
405 410 415

Ser Ala Asn Lys Ala
420

<210> 104
<211> 449
<212> PRT
<213> Trichomonas vaginalis

<400> 104

Met Leu Ala Ser Ser Ser Arg Ala Ala Ala Asn Ile Arg Trp Val Asp
1 5 10 15

Thr Ser His Asn Ala Ile Ala Phe Asp Met His Lys Cys Ile Asn Cys
20 25 30

Gln Ala Cys Val Arg Ala Cys Lys Asn Val Ala Gly Gln Ser Val Leu
35 40 45

Lys Ser Val Lys Ile Asn Glu Gly Lys Lys Lys Gly Val Val Gln Thr
50 55 60

Val Thr Gly Lys Leu Leu Ala Glu Thr Asn Cys Ile Gly Cys Gly Gln
65 70 75 80

Cys Thr Leu Val Cys Pro Thr Gln Ala Ile His Glu Lys Asp Ala Leu
85 90 95

Lys Gln Met Asn Asn Ile Phe Lys Asn Lys Gly Asp Arg Ile Leu Val
100 105 110

Cys Gln Ile Ala Pro Ala Ile Arg Ile Asn Met Arg Arg Pro Trp Cys
115 120 125

Ser Ser Arg Asn Ser Phe His Arg Gln Ser Arg Tyr Ser Pro Gln Arg
130 135 140

Leu Gly Phe Asp Tyr Val Phe Asp Thr Asn Phe Gly Ala Asp Leu Thr
145 150 155 160

Ile Val Glu Glu Ala Thr Glu Leu Leu Gln Arg Leu Asn Asp Pro Lys
165 170 175

Ala Val Leu Pro Met Phe Thr Ser Cys Cys Pro Ala Trp Val Asn Tyr
180 185 190

Val Glu Lys Ser Tyr Pro Gln Trp Met Pro His Leu Ser Thr Cys Arg
195 200 205

Ser Pro Ile Gly Met Leu Ser Ala Val Ile Lys Asn Val Phe Pro Lys
210 215 220

His Ile Gly Val Asp Pro Lys Arg Ile Phe Ser Val Gly Ile Met Pro
225 230 235 240

Cys Thr Ala Lys Lys Asp Glu Ala Ala Arg Glu Gln Leu Met Thr Lys
245 250 255

Ser Gly Leu His Glu Thr Asp Leu Asp Ile Thr Ser Arg Glu Leu Ala
260 265 270

Lys Met Ile Lys Ala Ala Lys Ile Asn Phe Lys Glu Leu Pro Asp Thr
275 280 285

Glu Leu Asp Ser Pro Tyr Ala Met Ala Thr Gly Gly Gly Ala Ile Phe
290 295 300

Cys Ala Thr Gly Gly Val Met Glu Ala Ala Val Arg Ser Ala Tyr Lys
305 310 315 320

Phe Ala Thr Gly Lys Glu Leu Ala Pro Ile Glu Phe Val Gln Val Arg
325 330 335

Gly Ala Glu Lys Gly Ile Lys Val Gly Thr Val Asp Ile Asn Gly Arg
340 345 350

Glu Ile Lys Val Ala Val Ala Gln Gly Val Lys Asn Ala Met Ser Leu
355 360 365

Ile Lys Lys Ile Glu Glu Gly Gln Asp Asp Val Lys Gly Val Val Phe
370 375 380

Cys Glu Val Met Ala Cys Pro Gly Gly Cys Val Gly Gly Gly Gly Ser
385 390 395 400

Pro Arg Ala Lys Thr Lys Ala Ala Met Asn Lys Arg Leu Asp Ala Thr
405 410 415

Tyr Arg Ile Asp Arg Ala Ser Lys Tyr Arg Thr Pro Gln Asp Asn Thr
420 425 430

Gln Leu Gln Asp Leu Tyr Asn Ala Thr Trp Val Val Ser Leu Val Met
435 440 445

Asp

<210> 105
<211> 645
<212> PRT
<213> T. maritima

<400> 105

Met Lys Ile Tyr Val Asp Gly Arg Glu Val Ile Ile Asn Asp Asn Glu
1 5 10 15

Arg Asn Leu Leu Glu Ala Leu Lys Asn Val Gly Ile Glu Ile Pro Asn
20 25 30

Leu Cys Tyr Leu Ser Glu Ala Ser Ile Tyr Gly Ala Cys Arg Met Cys
35 40 45

Leu Val Glu Ile Asn Gly Gln Ile Thr Thr Ser Cys Thr Leu Lys Pro
50 55 60

Tyr Glu Gly Met Lys Val Lys Thr Asn Thr Pro Glu Ile Tyr Glu Met
65 70 75 80

Arg Arg Asn Ile Leu Glu Leu Ile Leu Ala Thr His Asn Arg Asp Cys
85 90 95

Thr Thr Cys Asp Arg Asn Gly Ser Cys Lys Leu Gln Lys Tyr Ala Glu
100 105 110

Asp Phe Gly Ile Arg Lys Ile Arg Phe Glu Ala Leu Lys Lys Glu His
115 120 125

Val Arg Asp Glu Ser Ala Pro Val Val Arg Asp Thr Ser Lys Cys Ile
130 135 140

Leu Cys Gly Asp Cys Val Arg Val Cys Glu Glu Ile Gln Gly Val Gly
145 150 155 160

Val Ile Glu Phe Ala Lys Arg Gly Phe Glu Ser Val Val Thr Thr Ala
165 170 175

Phe Asp Thr Pro Leu Ile Glu Thr Glu Cys Val Leu Cys Gly Gln Cys
180 185 190

Val Ala Tyr Cys Pro Thr Gly Ala Leu Ser Ile Arg Asn Asp Ile Asp
195 200 205

Lys Leu Ile Glu Ala Leu Glu Ser Asp Lys Ile Val Ile Gly Met Ile
210 215 220

Ala Pro Ala Val Arg Ala Ala Ile Gln Glu Glu Phe Gly Ile Asp Glu
225 230 235 240

Asp Val Ala Met Ala Glu Lys Leu Val Ser Phe Leu Lys Thr Ile Gly
245 250 255

Phe Asp Lys Val Phe Asp Val Ser Phe Gly Ala Asp Leu Val Ala Tyr
260 265 270

Glu Glu Ala His Glu Phe Tyr Glu Arg Leu Lys Lys Gly Glu Arg Leu
275 280 285

Pro Gln Phe Thr Ser Cys Cys Pro Ala Trp Val Lys His Ala Glu His
290 295 300

Thr Tyr Pro Gln Tyr Leu Gln Asn Leu Ser Ser Val Lys Ser Pro Gln
305 310 315 320

Gln Ala Leu Gly Thr Val Ile Lys Lys Ile Tyr Ala Arg Lys Leu Gly
325 330 335

Val Pro Glu Glu Lys Ile Phe Leu Val Ser Phe Met Pro Cys Thr Ala
340 345 350

Lys Lys Phe Glu Ala Glu Arg Glu Glu His Glu Gly Ile Val Asp Ile
355 360 365

Val Leu Thr Thr Arg Glu Leu Ala Gln Leu Ile Lys Met Ser Arg Ile
370 375 380

Asp Ile Asn Arg Val Glu Pro Gln Pro Phe Asp Arg Pro Tyr Gly Val
385 390 395 400

Ser Ser Gln Ala Gly Leu Gly Phe Gly Lys Ala Gly Gly Val Phe Ser
405 410 415

Cys Val Leu Ser Val Leu Asn Glu Glu Ile Gly Ile Glu Lys Val Asp
420 425 430

Val Lys Ser Pro Glu Asp Gly Ile Arg Val Ala Glu Val Thr Leu Lys
435 440 445

Asp Gly Thr Ser Phe Lys Gly Ala Val Ile Tyr Gly Leu Gly Lys Val
450 455 460

Lys Lys Phe Leu Glu Glu Arg Lys Asp Val Glu Ile Ile Glu Val Met
465 470 475 480

Ala Cys Asn Tyr Gly Cys Val Gly Gly Gly Gly Gln Pro Tyr Pro Asn
485 490 495

Asp Ser Arg Ile Arg Glu His Arg Ala Lys Val Leu Arg Asp Thr Met
500 505 510

Gly Ile Lys Ser Leu Leu Thr Pro Val Glu Asn Leu Phe Leu Met Lys
515 520 525

Leu Tyr Glu Glu Asp Leu Lys Asp Glu His Thr Arg His Glu Ile Leu
530 535 540

His Thr Thr Tyr Arg Pro Arg Arg Arg Tyr Pro Glu Lys Asp Val Glu
545 550 555 560

Ile Leu Pro Val Pro Asn Gly Glu Lys Arg Thr Val Lys Val Cys Leu
565 570 575

Gly Thr Ser Cys Tyr Thr Lys Gly Ser Tyr Glu Ile Leu Lys Lys Leu
580 585 590

Val Asp Tyr Val Lys Glu Asn Asp Met Glu Gly Lys Ile Glu Val Leu
595 600 605

Gly Thr Phe Cys Val Glu Asn Cys Gly Ala Ser Pro Asn Val Ile Val
610 615 620

Asp Asp Lys Ile Ile Gly Gly Ala Thr Phe Glu Lys Val Leu Glu Glu
625 630 635 640

Leu Ser Lys Asn Gly

<210> 106
<211> 369
<212> PRT
<213> T vaginalis

<400> 106

Cys Asp Gly Lys Trp Leu Ala Pro Ala Cys Val Thr Thr Val Trp Asp
1 5 10 15

Gly Leu Lys Ile Asp Thr Lys Ser Lys Met Val Lys Glu Ser Val Glu
20 25 30

Asn Asn Leu Lys Glu Leu Leu Asp Cys His Asp Glu Thr Cys Ser Ser
35 40 45

Cys Val Ala Asn His Arg Cys Gln Phe Arg Asp Met Asn Val Ala Tyr
50 55 60

Ser Ile Lys Ala Glu Thr Lys Glu Glu Cys Ser Glu Glu Gly Ile Asp
65 70 75 80

Glu Ser Thr Asn Ser Ile Arg Leu Asp Thr Ser Lys Cys Val Leu Cys
85 90 95

Gly Arg Cys Ile Arg Ala Cys Glu Glu Val Ala Gly Gln Ser Ala Ile
100 105 110

Ile Phe Gly Asn Arg Ala Lys His Met Arg Ile Gln Pro Thr Phe Gly
115 120 125

Gln Thr Leu Gln Asp Thr Ser Cys Ile Lys Cys Gly Gln Cys Thr Leu
130 135 140

Tyr Cys Pro Val Gly Ala Ile Thr Glu Lys Ser Gln Val Lys Gln Ala
145 150 155 160

Leu Asp Ile Leu Ser Asn Lys Gly Lys Lys Ile Ser Val Ile Gln Val
165 170 175

Ala Pro Ala Val Arg Val Ala Leu Ser Glu Ala Phe Gly Tyr Lys Glu
180 185 190

Gly Ser Val Thr Thr Gly Lys Met Val Ser Ala Leu Lys Ala Leu Gly
195 200 205

Phe Asp Tyr Val Tyr Asp Thr Asn Tyr Ser Ala Asp Leu Thr Ile Val
210 215 220

Glu Glu Ala Gly Glu Leu Val Gln Arg Leu Lys Asn Pro Asn Ala Val
225 230 235 240

Phe Pro Met Phe Thr Ser Cys Cys Pro Ala Trp Val Asn Tyr Val Glu
245 250 255

Gln Ser Ala Pro Asp Phe Ile Pro Asn Leu Ser Ser Cys Arg Ser Pro

260 265 270

Gln Gly Met Leu Ser Ser Leu Val Lys Asn Tyr Leu Pro Lys Val Leu
275 280 285

Asn Ile Pro Val Glu Asp Val Leu Asn Phe Ser Ile Met Pro Cys Thr
290 295 300

Ala Lys Lys Asp Glu Ile Glu Arg Pro Glu Leu Arg Thr Lys Asp Gly
305 310 315 320

His Lys Glu Thr Asp Met Val Leu Thr Val Arg Glu Leu Val Glu Met
325 330 335

Ile Lys Leu Ser Gly Ile Asp Phe Asn Asn Leu Pro Asp Thr Pro Phe
340 345 350

Asp Ser Ile Phe Gly Phe Gly Ser Gly Ala Gly Gln Ile Phe Ala Ala
355 360 365

Thr

<210> 107
<211> 476
<212> PRT
<213> R. norvegicus

<400> 107

Met Ala Ser Pro Phe Ser Gly Ala Leu Gln Leu Thr Asp Leu Asp Asp
1 5 10 15

Phe Ile Gly Pro Ser Gln Ser Cys Ile Lys Pro Val Thr Val Ala Lys
20 25 30

Lys Pro Gly Ser Gly Ile Ala Lys Ile His Ile Glu Asp Asp Gly Ser
35 40 45

Tyr Phe Gln Val Asn Pro Asp Gly Arg Ser Gln Lys Leu Glu Lys Ala
50 55 60

Lys Val Ser Leu Asn Asp Cys Leu Ala Cys Ser Gly Cys Val Thr Ser
65 70 75 80

Ala Glu Thr Ile Leu Ile Thr Gln Gln Ser His Glu Glu Leu Arg Lys
85 90 95

Val Leu Asp Ala Asn Lys Val Ala Ala Pro Gly Gln Gln Arg Leu Val
100 105 110

Val Val Ser Val Ser Pro Gln Ser Arg Ala Ser Leu Ala Ala Arg Phe
115 120 125

Gln Leu Asp Ser Thr Asp Thr Ala Arg Lys Leu Thr Ser Phe Phe Lys
130 135 140

Lys Ile Gly Val His Phe Val Phe Asp Thr Ala Phe Ala Arg Asn Phe

145 150 155 160

Ser Leu Leu Glu Ser Gln Lys Glu Phe Val Gln Arg Phe Arg Glu Gln
165 170 175

Ala Asn Ser Arg Glu Ala Leu Pro Met Leu Ala Ser Ala Cys Pro Gly
180 185 190

Trp Ile Cys Tyr Ala Glu Lys Thr His Gly Asn Phe Ile Leu Pro Tyr
195 200 205

Ile Ser Thr Ala Arg Ser Pro Gln Gln Val Met Gly Ser Leu Ile Lys
210 215 220

Asp Phe Phe Ala Gln Gln Gln Leu Leu Thr Pro Asp Lys Ile Tyr His
225 230 235 240

Val Thr Val Met Pro Cys Tyr Asp Lys Lys Leu Glu Ala Ser Arg Pro
245 250 255

Asp Phe Phe Asn Gln Glu Tyr Gln Thr Arg Asp Val Asp Cys Val Leu
260 265 270

Thr Thr Gly Glu Val Phe Arg Leu Leu Glu Glu Glu Gly Val Ser Leu
275 280 285

Ser Glu Leu Glu Pro Val Pro Leu Asp Gly Leu Thr Arg Ser Val Ser
290 295 300

Ala Glu Glu Pro Thr Ser His Arg Gly Gly Gly Ser Gly Gly Tyr Leu
305 310 315 320

Glu His Val Phe Arg His Ala Ala Gln Glu Leu Phe Gly Ile His Val
325 330 335

Ala Asp Val Thr Tyr Gln Pro Met Arg Asn Lys Asp Phe Gln Glu Val
340 345 350

Thr Leu Glu Arg Glu Gly Gln Val Leu Leu Arg Phe Ala Val Ala Tyr
355 360 365

Gly Phe Arg Asn Ile Gln Asn Leu Val Gln Lys Leu Lys Arg Gly Arg
370 375 380

Cys Pro Tyr His Tyr Val Glu Val Met Ala Cys Pro Ser Gly Cys Leu
385 390 395 400

Asn Gly Gly Gly Gln Leu Lys Ala Pro Asp Thr Glu Gly Arg Glu Leu
405 410 415

Leu Gln Gln Val Glu Arg Leu Tyr Ser Met Val Arg Thr Glu Ala Pro
420 425 430

Glu Asp Ala Pro Gly Val Gln Glu Leu Tyr Gln His Trp Leu Gln Gly
435 440 445

Glu Asp Ser Glu Arg Ala Ser His Leu Leu His Thr Gln Tyr His Ala
450 455 460

Val Glu Lys Ile Asn Ser Gly Leu Ser Ile Arg Trp
465 470 475

<210> 108
<211> 525
<212> PRT
<213> S. cerevisiae

<400> 108

Met Ala Ser Pro Phe Ser Gly Ala Leu Gln Leu Thr Asp Leu Asp Asp
1 5 10 15

Phe Ile Gly Pro Ser Gln Val Gly Ser Leu Gln Ala Leu Leu Ala Leu
20 25 30

Ala Phe Leu His Thr Gly Asn Phe Ser Ala Ala Gly Cys Trp Glu Pro
35 40 45

Asp Pro Trp Glu Cys Ile Lys Pro Val Lys Val Glu Lys Arg Ala Gly
50 55 60

Ser Gly Val Ala Lys Ile Arg Ile Glu Asp Asp Gly Ser Tyr Phe Gln
65 70 75 80

Ile Asn Gln Glu Lys Leu Gly Glu Leu Glu Leu Glu Pro Thr Phe Gly
85 90 95

Ile Phe Leu Pro Tyr Ser Pro Asp Gly Gly Thr Arg Arg Leu Glu Lys
100 105 110

Ala Lys Val Ser Leu Asn Asp Cys Leu Ala Cys Ser Gly Cys Ile Thr
115 120 125

Ser Ala Glu Thr Val Leu Ile Thr Gln Gln Ser His Glu Glu Leu Lys
130 135 140

Lys Val Leu Asp Ala Asn Lys Met Ala Ala Pro Ser Gln Gln Arg Leu
145 150 155 160

Val Val Val Ser Val Ser Pro Gln Ser Arg Ala Ser Leu Ala Ala Arg
165 170 175

Phe Gln Leu Asn Pro Thr Asp Thr Ala Arg Lys Leu Thr Ser Phe Phe
180 185 190

Lys Lys Ile Gly Val His Phe Val Phe Asp Thr Ala Phe Ser Arg His
195 200 205

Phe Ser Leu Leu Glu Ser Gln Arg Glu Phe Val Arg Arg Phe Arg Gly
210 215 220

Gln Ala Asp Cys Arg Gln Ala Leu Pro Leu Leu Ala Ser Ala Cys Pro
225 230 235 240

Gly Trp Ile Cys Tyr Ala Glu Lys Thr His Gly Ser Phe Ile Leu Pro
245 250 255

His Ile Ser Thr Ala Arg Ser Pro Gln Gln Val Met Gly Ser Leu Val
260 265 270

Lys Asp Phe Phe Ala Gln Gln Gln His Leu Thr Pro Asp Lys Ile Tyr
275 280 285

His Val Thr Val Met Pro Cys Tyr Asp Lys Lys Leu Glu Ala Ser Arg
290 295 300

Pro Asp Phe Phe Asn Gln Glu His Gln Thr Arg Asp Val Asp Cys Val
305 310 315 320

Leu Thr Thr Gly Glu Val Phe Arg Leu Leu Glu Glu Glu Gly Val Ser
325 330 335

Leu Pro Asp Leu Glu Pro Ala Pro Leu Asp Ser Leu Cys Ser Gly Ala
340 345 350

Ser Ala Glu Glu Pro Thr Ser His Arg Gly Gly Gly Ser Gly Gly Tyr
355 360 365

Leu Glu His Val Phe Arg His Ala Ala Arg Glu Leu Phe Gly Ile His
370 375 380

Val Ala Glu Val Thr Tyr Lys Pro Leu Arg Asn Lys Asp Phe Gln Glu
385 390 395 400

Val Thr Leu Glu Lys Glu Gly Gln Val Leu Leu His Phe Ala Met Ala
405 410 415

Tyr Gly Phe Arg Asn Ile Gln Asn Leu Val Gln Arg Leu Lys Arg Gly
420 425 430

Arg Cys Pro Tyr His Tyr Val Glu Val Met Ala Cys Pro Ser Gly Cys
435 440 445

Leu Asn Gly Gly Gly Gln Leu Gln Ala Pro Asp Arg Pro Ser Arg Glu
450 455 460

Leu Leu Gln His Val Glu Arg Leu Tyr Gly Met Val Arg Ala Glu Ala
465 470 475 480

Pro Glu Asp Ala Pro Gly Val Gln Glu Leu Tyr Thr His Trp Leu Gln
485 490 495

Gly Thr Asp Ser Glu Cys Ala Gly Arg Leu Leu His Thr Gln Tyr His
500 505 510

Ala Val Glu Lys Ala Ser Thr Gly Leu Gly Ile Arg Trp
515 520 525

<210> 109
<211> 572
<212> PRT

<213> C. perfringens

<400> 109

Met Asn Lys Ile Ile Ile Asn Asp Lys Thr Ile Glu Phe Asp Gly Asp
1 5 10 15

Lys Thr Ile Leu Asp Leu Ala Arg Glu Asn Gly Phe Asp Ile Pro Val
20 25 30

Leu Cys Glu Leu Lys Asn Cys Gly Asn Lys Gly Gln Cys Gly Val Cys
35 40 45

Leu Val Glu Gln Glu Gly Asn Asp Arg Leu Leu Arg Ser Cys Ala Ile
50 55 60

Lys Ala Lys Asp Gly Met Val Ile Lys Thr Asp Ser Glu Lys Val Leu
65 70 75 80

Glu Ala Arg Lys Glu Arg Val Ala Glu Leu Leu Asp Glu His Glu Phe
85 90 95

Lys Cys Gly Pro Cys Lys Arg Arg Glu Asn Cys Glu Phe Leu Lys Leu
100 105 110

Val Ile Lys Thr Lys Ala Arg Ala His Lys Pro Phe Val Val Ala Asp
115 120 125

Lys Ser Glu Tyr Val Asp Asp Arg Ser Lys Ser Ile Val Leu Asp Arg
130 135 140

Ser Lys Cys Val Lys Cys Gly Arg Cys Val Ala Ala Cys Arg Thr Arg
145 150 155 160

Thr Ala Thr Asn Ser Ile Lys Phe His Arg Ile Asp Gly Val Arg Leu
165 170 175

Val Gly Pro Glu Glu Leu Lys Cys Phe Asp Asp Thr Asn Cys Leu Leu
180 185 190

Cys Gly Gln Cys Ile Ala Ala Cys Pro Val Asp Ala Leu Ser Glu Lys
195 200 205

Ser His Ile Glu Arg Val Gln Asp Ala Leu Asn Asp Pro Glu Lys His
210 215 220

Val Ile Val Ala Met Ala Pro Ala Val Arg Thr Ser Met Gly Glu Leu
225 230 235 240

Phe Lys Met Gly Tyr Gly Gln Asp Val Thr Gly Lys Leu Tyr Thr Ala
245 250 255

Leu Arg Glu Leu Gly Phe Asp Lys Val Phe Asp Ile Asn Phe Gly Ala
260 265 270

Asp Met Thr Ile Met Glu Glu Ala Thr Glu Leu Ile Glu Arg Ile Lys
275 280 285

```
Asn Asn Gly Pro Phe Pro Met Leu Thr Ser Cys Cys Pro Ser Trp Val
    290                 295                 300

Arg Glu Val Glu Asn Tyr Phe Pro Glu Leu Val Glu Asn Leu Ser Ser
305                 310                 315                 320

Ala Lys Ser Pro Gln Gln Ile Phe Gly Ala Ala Ser Lys Thr Tyr Tyr
                325                 330                 335

Pro Gln Val Ala Asp Ile Asp Pro Lys Lys Val Phe Thr Val Thr Val
            340                 345                 350

Met Pro Cys Thr Ser Lys Lys Phe Glu Ala Asp Arg Pro Glu Met Glu
        355                 360                 365

Asn Glu Gly Ile Arg Asn Ile Asp Ala Val Ile Thr Thr Arg Glu Leu
    370                 375                 380

Ala Arg Met Ile Lys Ala Ala Lys Ile Asp Phe Ala Lys Leu Glu Asp
385                 390                 395                 400

Gly Glu Val Asp Pro Ala Met Gly Glu Tyr Thr Gly Ala Gly Val Ile
                405                 410                 415

Phe Gly Ala Thr Gly Gly Val Met Glu Ala Ala Leu Arg Thr Ala Lys
            420                 425                 430

Asp Phe Met Glu Asn Asp Asn Leu Asp Asn Val Asp Tyr Glu Ala Val
        435                 440                 445

Arg Gly Leu Ala Gly Ile Lys Glu Ala Glu Val Glu Ile Ala Gly Asn
    450                 455                 460

Glu Tyr Lys Leu Ala Val Val Ser Gly Ala Ala Asn Val Phe Glu Leu
465                 470                 475                 480

Val Lys Ser Gly Lys Ile Asn Asp Tyr His Phe Ile Glu Val Met Ala
                485                 490                 495

Cys Pro Gly Gly Cys Val Asn Gly Gly Gly Gln Pro His Ile Ser Ala
            500                 505                 510

Glu Asp Ser Asp Lys Ile Asp Ile Arg Glu Val Arg Ala Ser Val Leu
        515                 520                 525

Tyr Asn Gln Asp Lys Asn Leu Glu Lys Arg Lys Ser His Gln Asn Ser
    530                 535                 540

Ala Leu Leu Lys Met Tyr Glu Asn Tyr Met Gly Lys Pro Gly His Gly
545                 550                 555                 560

Arg Ala His Glu Leu Leu His Met Lys Tyr Lys Lys
                565                 570
```

<210> 110
<211> 572

<212> PRT
<213> C. perfringens

<400> 110

Met Asn Lys Ile Ile Ile Asn Asp Lys Thr Ile Glu Phe Asp Gly Asp
1 5 10 15

Lys Thr Ile Leu Asp Leu Ala Arg Glu Asn Gly Phe Asp Ile Pro Val
20 25 30

Leu Cys Glu Leu Lys Asn Cys Gly Asn Lys Gly Gln Cys Gly Val Cys
35 40 45

Leu Val Glu Gln Glu Gly Asn Asp Arg Leu Leu Arg Ser Cys Ala Ile
50 55 60

Lys Ala Lys Asp Gly Met Val Ile Lys Thr Asp Ser Glu Lys Val Leu
65 70 75 80

Glu Ala Arg Lys Glu Arg Val Ala Glu Leu Leu Asp Glu His Glu Phe
85 90 95

Lys Cys Gly Pro Cys Lys Arg Arg Glu Asn Cys Glu Phe Leu Lys Leu
100 105 110

Val Ile Lys Thr Lys Ala Arg Ala His Lys Pro Phe Val Val Ala Asp
115 120 125

Lys Ser Glu Tyr Val Asp Asp Arg Ser Lys Ser Ile Val Leu Asp Arg
130 135 140

Ser Lys Cys Val Lys Cys Gly Arg Cys Val Ala Ala Cys Arg Thr Arg
145 150 155 160

Thr Ala Thr Asn Ser Ile Lys Phe His Arg Ile Asp Gly Val Arg Leu
165 170 175

Val Gly Pro Glu Glu Leu Lys Cys Phe Asp Asp Thr Asn Cys Leu Leu
180 185 190

Cys Gly Gln Cys Ile Ala Ala Cys Pro Val Asp Ala Leu Ser Glu Lys
195 200 205

Ser His Ile Glu Arg Val Gln Glu Ala Leu Asn Asp Pro Glu Lys His
210 215 220

Val Ile Val Ala Met Ala Pro Ala Val Arg Thr Ser Met Gly Glu Leu
225 230 235 240

Phe Lys Met Gly Tyr Gly Gln Asp Val Thr Gly Lys Leu Tyr Thr Ala
245 250 255

Leu Arg Glu Leu Gly Phe Asp Lys Val Phe Asp Ile Asn Phe Gly Ala
260 265 270

Asp Met Thr Ile Met Glu Glu Ala Thr Glu Leu Ile Glu Arg Ile Lys
275 280 285

```
Asn Asn Gly Pro Phe Pro Met Leu Thr Ser Cys Cys Pro Ser Trp Val
    290                 295                 300

Arg Glu Val Glu Asn Tyr Phe Pro Glu Leu Val Glu Asn Leu Ser Ser
305                 310                 315                 320

Ala Lys Ser Pro Gln Gln Ile Phe Gly Ala Ala Ser Lys Thr Tyr Tyr
                325                 330                 335

Pro Gln Val Ala Asp Ile Asp Pro Lys Lys Val Phe Thr Val Thr Val
            340                 345                 350

Met Pro Cys Thr Ser Lys Lys Phe Glu Ala Asp Arg Pro Glu Met Glu
        355                 360                 365

Asn Glu Gly Ile Arg Asn Ile Asp Ala Val Ile Thr Thr Arg Glu Leu
    370                 375                 380

Ala Arg Met Ile Lys Ala Ala Lys Ile Asp Phe Ala Lys Leu Glu Asp
385                 390                 395                 400

Gly Glu Val Asp Pro Ala Met Gly Glu Tyr Thr Gly Ala Gly Val Ile
                405                 410                 415

Phe Gly Ala Thr Gly Gly Val Met Glu Ala Ala Leu Arg Thr Ala Lys
            420                 425                 430

Asp Phe Met Glu Asn Asp Asn Leu Asp Asn Val Asp Tyr Glu Ala Val
        435                 440                 445

Arg Gly Leu Ala Gly Ile Lys Glu Ala Glu Val Glu Ile Ala Gly Asn
    450                 455                 460

Glu Tyr Lys Leu Ala Val Val Ser Gly Ala Ala Asn Val Phe Glu Leu
465                 470                 475                 480

Val Lys Ser Gly Lys Ile Asn Asp Tyr His Phe Ile Glu Val Met Ala
                485                 490                 495

Cys Pro Gly Gly Cys Val Asn Gly Gly Gly Gln Pro His Ile Ser Ala
            500                 505                 510

Glu Asp Ser Asp Lys Met Asp Ile Arg Glu Val Arg Ala Ser Val Leu
        515                 520                 525

Tyr Asn Gln Asp Lys Asn Leu Glu Lys Arg Lys Ser His Gln Asn Ser
    530                 535                 540

Ala Leu Leu Lys Met Tyr Glu Ser Tyr Met Gly Lys Pro Gly His Gly
545                 550                 555                 560

Arg Ala His Glu Leu Leu His Met Lys Tyr Lys Lys
                565                 570
```

<210> 111

<211> 494
<212> PRT
<213> C. tetani

<400> 111

Met Ile Val Phe Glu Asn Gln Leu Lys Lys Leu Lys Tyr Leu Val Leu
1 5 10 15

Lys Glu Val Ala Lys Met Thr Leu Glu Asp Arg Leu Gly Glu Glu Asp
20 25 30

Ile Glu Arg Ile Ser Phe Asp Ile Ile Lys Gly Asp Lys Ala Glu Tyr
35 40 45

Arg Cys Cys Val Tyr Lys Glu Arg Ala Ile Val Tyr Glu Arg Ala Lys
50 55 60

Leu Ala Thr Gly Cys Leu Pro Asn Gly Gln Val Ala Glu Glu Phe Val
65 70 75 80

His Val Glu Asp Asp Asp Gln Ile Ile Tyr Val Ile Asp Ala Ala Cys
85 90 95

Asp Lys Cys Pro Ile Asn Lys Tyr Val Val Thr Glu Ala Cys Arg Gly
100 105 110

Cys Leu Gln His Lys Cys Met Glu Val Cys Pro Ala Gly Ser Ile Asn
115 120 125

Arg Ala Ala Gly Lys Ala Tyr Ile Asn His Glu Thr Cys Lys Glu Cys
130 135 140

Gly Leu Cys Glu Ser Ala Cys Pro Tyr Asn Ala Ile Ala Glu Val Met
145 150 155 160

Arg Pro Cys Arg Arg Ala Cys Pro Thr Gly Ala Leu Gln Met Asn Leu
165 170 175

Glu Asp Asn Lys Ala Thr Ile Asn Lys Glu Asp Cys Ile Asn Cys Gly
180 185 190

Ser Cys Met Ser Val Cys Pro Phe Gly Ala Ile Ser Asp Lys Ser Tyr
195 200 205

Ile Val Asp Ile Thr Lys Ala Leu Lys Asn Asn Lys Lys Val Tyr Ala
210 215 220

Met Val Ala Pro Ala Ile Thr Gly Gln Phe Gly Lys Asp Val Ser Val
225 230 235 240

Gly Lys Met Lys Asn Ala Phe Lys Ala Met Gly Phe Glu Asp Met Leu
245 250 255

Glu Val Ala Cys Gly Ala Asp Ala Val Ala Ala His Glu Ser Glu Glu
260 265 270

Phe Ile Glu Arg Leu Glu Ser Gly Lys Lys Tyr Met Thr Thr Ser Cys

275 280 285

Cys Pro Gly Phe Leu Gly Tyr Ile Glu Lys Lys Phe Pro Asp Gln Leu
290 295 300

Glu Asn Val Ser Asn Thr Val Ser Pro Met Val Ala Ile Gly Arg Met
305 310 315 320

Ile Lys Lys Glu Tyr Glu Asp Ser Val Val Val Phe Val Gly Pro Cys
325 330 335

Thr Ala Lys Lys Ala Glu Ile Lys Arg Lys Gly Ile Lys Asp Ala Val
340 345 350

Asp Tyr Val Met Thr Phe Glu Glu Ile Ala Ala Leu Met Gly Ala Phe
355 360 365

Glu Ile Asp Pro Ala Glu Cys Glu Glu Glu Asp Ile Asn Asp Gly Ser
370 375 380

Asn Tyr Gly Arg Gly Phe Ala Gln Gly Gly Gly Val Val Ser Ala Ile
385 390 395 400

Gln Asn Cys Ile Lys Asp Lys Glu Gly Ile Lys Phe Asn Pro Leu Arg
405 410 415

Val Ser Gly Pro Asp Gln Ile Lys Arg Ala Met Ile Met Ala Lys Val
420 425 430

Gly Lys Leu Ser Glu Asn Phe Ile Glu Gly Met Met Cys Glu Gly Gly
435 440 445

Cys Ile Gly Gly Pro Ala Thr Met Val Ser Ala Val Lys Ala Lys Ala
450 455 460

Pro Leu Met Lys Phe Ser Lys Ser Ser Thr Ile Lys Asp Val Lys Asp
465 470 475 480

Asn Glu Val Leu Asp Lys Tyr Lys Asp Ile Asn Met Glu Arg
485 490

<210> 112
<211> 448
<212> PRT
<213> C. tetani

<400> 112

Met His Asn Asp Tyr Arg Glu Ile Phe Lys Arg Leu Ser Lys Ser Tyr
1 5 10 15

Tyr Asp Asp Thr Phe Glu Lys Glu Val Glu Asn Ile Leu Ser Ser His
20 25 30

Ser Met Asp Arg Glu Lys Leu Ala Lys Ile Ile Ser Ile Leu Cys Gly
35 40 45

Val Asn Ile Glu His Ser Glu Asn Tyr Ile Ser Asn Leu Lys Asn Ala

Ile Lys Asn Tyr Thr Ala Ser Ala Glu Lys Val Val Thr Lys Leu Pro
65 70 75 80

Cys Ser Thr Gln Cys Ala Lys Asp Gly Asp Ile Ile Cys Glu Lys Ser
85 90 95

Cys Pro Val Asn Ala Ile Phe Arg Asp Pro Asn Asp Asn Asn Ile Tyr
100 105 110

Ile Asn Asp Glu Leu Cys Leu Asp Cys Gly Leu Cys Val Arg Asn Cys
115 120 125

Pro Ser Gly Ser Ile Leu Asp Lys Lys Glu Phe Ile Pro Leu Ala Glu
130 135 140

Leu Leu Lys Ser Glu Ser Ile Val Ile Ala Ala Val Ala Pro Ala Ile
145 150 155 160

Met Gly Gln Phe Gly Glu Asn Thr Thr Ile Asn Gln Leu Arg Thr Ala
165 170 175

Phe Lys Lys Leu Gly Phe Thr Asp Met Val Glu Val Ala Phe Phe Ala
180 185 190

Asp Met Leu Thr Leu Lys Glu Ala Val Glu Tyr Asp His Phe Val Lys
195 200 205

Asp Glu Gln Asp Phe Met Ile Thr Ser Cys Cys Cys Pro Met Trp Val
210 215 220

Gly Met Leu Lys Lys Val Tyr Asn Asp Leu Val Lys Tyr Val Ser Pro
225 230 235 240

Ser Val Ser Pro Met Ile Ala Ala Gly Arg Val Leu Lys Leu Leu Asn
245 250 255

Pro Asn Cys Lys Val Val Phe Val Gly Pro Cys Ile Ala Lys Lys Ala
260 265 270

Glu Ala Arg Glu Lys Asp Leu Leu Gly Asp Ile Asp Phe Val Leu Thr
275 280 285

Phe Thr Glu Leu Arg Asp Ile Phe Asp Val Phe Asp Ile Gln Pro Glu
290 295 300

Asn Leu Glu Glu Asp Phe Ser Ser Glu Tyr Ala Ser Lys Gly Gly Arg
305 310 315 320

Leu Tyr Ala Arg Thr Gly Gly Val Ser Ile Ala Val Ser Glu Ala Ile
325 330 335

Glu Lys Leu Phe Pro Asn Lys Tyr Lys Phe Leu Lys Thr Ile Gln Ala
340 345 350

Asp Gly Val Lys Gly Cys Lys Ser Leu Leu Asp Lys Ile Lys Gln Glu
355 360 365

Asp Ile Ser Ala Asn Phe Val Glu Gly Met Gly Cys Val Gly Gly Cys
370 375 380

Val Gly Gly Pro Lys Val Ile Ile Asp Pro Ser Glu Gly Arg Asn Ala
385 390 395 400

Val Asn Asn Phe Ala Glu Asn Ser Ser Ile Lys Val Ser Val Asp Ser
405 410 415

Asn Cys Met Asn Asp Ile Leu Ser Lys Ile Asn Ile Asn Ser Val Glu
420 425 430

Asp Phe Lys Asp Lys Asp Lys Ile Ser Ile Phe Glu Arg Glu Phe Lys
435 440 445

<210> 113
<211> 261
<212> PRT
<213> Pyrococcus furiosus

<400> 113

Met Gly Lys Val Arg Ile Gly Phe Tyr Ala Leu Thr Ser Cys Tyr Gly
1 5 10 15

Cys Gln Leu Gln Leu Ala Met Met Asp Glu Leu Leu Gln Leu Ile Pro
20 25 30

Asn Ala Glu Ile Val Cys Trp Phe Met Ile Asp Arg Asp Ser Ile Glu
35 40 45

Asp Glu Lys Val Asp Ile Ala Phe Ile Glu Gly Ser Val Ser Thr Glu
50 55 60

Glu Glu Val Glu Leu Val Lys Lys Ile Arg Glu Asn Ala Lys Ile Val
65 70 75 80

Val Ala Val Gly Ala Cys Ala Val Gln Gly Gly Val Gln Ser Trp Ser
85 90 95

Glu Lys Pro Leu Glu Glu Leu Trp Lys Lys Val Tyr Gly Asp Ala Lys
100 105 110

Val Lys Phe Gln Pro Lys Lys Ala Glu Pro Val Ser Lys Tyr Ile Lys
115 120 125

Val Asp Tyr Asn Ile Tyr Gly Cys Pro Pro Glu Lys Lys Asp Phe Leu
130 135 140

Tyr Ala Leu Gly Thr Phe Leu Ile Gly Ser Trp Pro Glu Asp Ile Asp
145 150 155 160

Tyr Pro Val Cys Leu Glu Cys Arg Leu Asn Gly His Pro Cys Ile Leu
165 170 175

Leu Glu Lys Gly Glu Pro Cys Leu Gly Pro Val Thr Arg Ala Gly Cys
180 185 190

Asn Ala Arg Cys Pro Gly Phe Gly Val Ala Cys Ile Gly Cys Arg Gly
195 200 205

Ala Ile Gly Tyr Asp Val Ala Trp Phe Asp Ser Leu Ala Lys Val Phe
210 215 220

Lys Glu Lys Gly Met Thr Lys Glu Glu Ile Ile Glu Arg Met Lys Met
225 230 235 240

Phe Asn Gly His Asp Glu Arg Val Glu Lys Met Val Glu Lys Ile Phe
245 250 255

Ser Gly Gly Glu Gln
260

<210> 114
<211> 252
<212> PRT
<213> Escherichia coli

<400> 114

Met Ser Pro Val Leu Thr Gln His Val Ser Gln Pro Ile Thr Leu Asp
1 5 10 15

Glu Gln Thr Gln Lys Met Lys Arg His Leu Leu Gln Asp Ile Arg Arg
20 25 30

Ser Ala Tyr Val Tyr Arg Val Asp Cys Gly Gly Cys Asn Ala Cys Glu
35 40 45

Ile Glu Ile Phe Ala Ala Ile Thr Pro Val Phe Asp Ala Glu Arg Phe
50 55 60

Gly Ile Lys Val Val Ser Ser Pro Arg His Ala Asp Ile Leu Leu Phe
65 70 75 80

Thr Gly Ala Val Thr Arg Ala Met Arg Met Pro Ala Leu Arg Ala Tyr
85 90 95

Glu Ser Ala Pro Asp His Lys Ile Cys Val Ser Tyr Gly Ala Cys Gly
100 105 110

Val Gly Gly Gly Ile Phe His Asp Leu Tyr Ser Val Trp Gly Gly Ser
115 120 125

Asp Thr Ile Val Pro Ile Asp Val Trp Ile Pro Gly Cys Pro Pro Thr
130 135 140

Pro Ala Ala Thr Ile His Gly Phe Ala Val Ala Leu Gly Leu Leu Gln
145 150 155 160

Gln Lys Ile His Ala Val Asp Tyr Arg Asp Pro Thr Gly Val Thr Met
165 170 175

Gln Pro Leu Trp Pro Gln Ile Pro Pro Ser Gln Arg Ile Ala Ile Glu
180 185 190

Arg Glu Ala Arg Arg Leu Ala Gly Tyr Arg Gln Gly Arg Glu Ile Cys
195 200 205

Asp Arg Leu Leu Arg His Leu Ser Asp Asp Pro Thr Gly Asn Arg Val
210 215 220

Asn Thr Trp Leu Arg Asp Ala Asp Asp Pro Arg Leu Asn Ser Ile Val
225 230 235 240

Gln Gln Leu Phe Arg Val Leu Arg Gly Leu His Asp
245 250

<210> 115
<211> 236
<212> PRT
<213> Methanothermobacter thermautotrophicus

<400> 115

Met Ala Glu Glu Asn Ala Lys Pro Arg Ile Gly Tyr Ile His Leu Ser
1 5 10 15

Gly Cys Thr Gly Asp Ala Met Ser Leu Thr Glu Asn Tyr Asp Ile Leu
20 25 30

Ala Glu Leu Leu Thr Asn Met Val Asp Ile Val Tyr Gly Gln Thr Leu
35 40 45

Val Asp Leu Trp Glu Met Pro Glu Met Asp Leu Ala Leu Val Glu Gly
50 55 60

Ser Val Cys Leu Gln Asp Glu His Ser Leu His Glu Leu Lys Glu Leu
65 70 75 80

Arg Glu Lys Ala Lys Leu Val Cys Ala Phe Gly Ser Cys Ala Gln Thr
85 90 95

Gly Cys Phe Thr Arg Tyr Ser Arg Gly Gly Gln Gln Ala Gln Pro Ser
100 105 110

His Glu Ser Phe Val Pro Ile Ala Asp Leu Ile Asp Val Asp Leu Ala
115 120 125

Ile Pro Gly Cys Pro Pro Ser Pro Glu Ile Ile Ala Lys Ala Val Val
130 135 140

Ala Leu Leu Asn Asn Asp Met Glu Tyr Leu Gln Pro Met Leu Asp Leu
145 150 155 160

Ala Gly Tyr Thr Glu Ala Cys Gly Cys Asp Leu Gln Thr Lys Val Val
165 170 175

Asn Gln Gly Leu Cys Thr Gly Cys Gly Thr Cys Ala Met Ala Cys Gln
180 185 190

Thr Arg Ala Leu Asp Met Thr Asn Gly Arg Pro Glu Leu Asn Ser Asp
195 200 205

Arg Cys Ile Lys Cys Gly Ile Cys Tyr Val Gln Cys Pro Arg Ser Trp
210 215 220

Trp Pro Glu Glu Gln Ile Lys Lys Glu Leu Gly Leu
225 230 235

<210> 116
<211> 259
<212> PRT
<213> Methanosarcina barkeri

<400> 116

Met Ala Asn Lys Ile Lys Leu Gly His Val His Leu Ser Gly Cys Thr
1 5 10 15

Gly Cys Leu Val Ser Val Ala Asp Asn Tyr Gln Gly Phe Leu Lys Ile
20 25 30

Leu Asp Asp Tyr Ala Asp Leu Val Tyr Cys Leu Thr Leu Ala Asp Val
35 40 45

Arg His Ile Pro Glu Met Asp Val Ala Leu Val Glu Gly Ser Val Cys
50 55 60

Ile Gln Asp Arg Glu Ser Val Glu Asp Ile Lys Glu Thr Arg Lys Lys
65 70 75 80

Ser Arg Ile Val Val Ala Leu Gly Ser Cys Ala Ser Tyr Gly Asn Ile
85 90 95

Thr Arg Phe Cys Arg Gly Gly Gln His Asn His Pro Gln His Glu Ser
100 105 110

Tyr Leu Pro Ile Gly Asp Leu Ile Asp Val Asp Val Tyr Ile Pro Gly
115 120 125

Cys Pro Pro Ser Pro Glu Leu Ile Arg Asn Val Ala Ile Met Ala Tyr
130 135 140

Leu Leu Leu Glu Gly Asn Glu Glu Gln Lys Asp Leu Ala Gly Arg Tyr
145 150 155 160

Leu Lys Pro Leu Met Asp Leu Ala Lys Arg Gly Thr Thr Gly Cys Phe
165 170 175

Cys Asp Leu Met Asp Asp Val Ile Asn Gln Gly Leu Cys Ile Gly Cys
180 185 190

Gly Ile Cys Ala Ala Ser Cys Pro Val Arg Ala Ile Thr His Glu Phe
195 200 205

Gly Lys Pro Gln Gly Asp Leu Asn Leu Cys Ile Lys Cys Gly Ser Cys
210 215 220

Tyr Gly Ala Cys Pro Arg Ser Phe Phe Asn Pro Asp Val Ile Ser Glu
225 230 235 240

Phe Glu Ser Ile Asn Glu Ile Ile Ala Gly Ala Leu Lys Glu Gly Glu
245 250 255

Lys Asp Asp

<210> 117
<211> 142
<212> PRT
<213> Rhodospirillum rubrum

<400> 117

Met Asn Phe Leu Ser Arg Met Ser Lys Lys Ser Pro Trp Leu Tyr Arg
1 5 10 15

Ile Asn Ala Gly Ser Cys Asn Gly Cys Asp Val Glu Leu Ala Thr Thr
20 25 30

Ala Cys Ile Pro Arg Tyr Asp Val Glu Arg Leu Gly Cys Gln Tyr Cys
35 40 45

Gly Ser Pro Lys His Ala Asp Ile Val Leu Val Thr Gly Pro Leu Thr
50 55 60

Ala Arg Val Lys Asp Lys Val Leu Arg Val Tyr Glu Glu Ile Pro Asp
65 70 75 80

Pro Lys Val Thr Val Ala Ile Gly Val Cys Pro Ile Ser Gly Gly Val
85 90 95

Phe Arg Glu Ser Tyr Ser Ile Val Gly Pro Ile Asp Arg Tyr Leu Pro
100 105 110

Val Asp Val Asn Val Pro Gly Cys Pro Pro Arg Pro Gln Ala Ile Ile
115 120 125

Glu Gly Ile Ala Lys Ala Ile Glu Ile Trp Ala Gly Arg Ile
130 135 140

<210> 118
<211> 428
<212> PRT
<213> Pyrococcus furiosus

<400> 118

Met Lys Asn Leu Tyr Leu Pro Ile Thr Ile Asp His Ile Ala Arg Val
1 5 10 15

Glu Gly Lys Gly Gly Val Glu Ile Ile Ile Gly Asp Asp Gly Val Lys
20 25 30

Glu Val Lys Leu Asn Ile Ile Glu Gly Pro Arg Phe Phe Glu Ala Ile
35 40 45

Thr Ile Gly Lys Lys Leu Glu Glu Ala Leu Ala Ile Tyr Pro Arg Ile

Cys Ser Phe Cys Ser Ala Ala His Lys Leu Thr Ala Leu Glu Ala Ala
65 70 75 80

Glu Lys Ala Val Gly Phe Val Pro Arg Glu Glu Ile Gln Ala Leu Arg
85 90 95

Glu Val Leu Tyr Ile Gly Asp Met Ile Glu Ser His Ala Leu His Leu
100 105 110

Tyr Leu Leu Val Leu Pro Asp Tyr Arg Gly Tyr Ser Ser Pro Leu Lys
115 120 125

Met Val Asn Glu Tyr Lys Arg Glu Ile Glu Ile Ala Leu Lys Leu Lys
130 135 140

Asn Leu Gly Thr Trp Met Met Asp Ile Leu Gly Ser Arg Ala Ile His
145 150 155 160

Gln Glu Asn Ala Val Leu Gly Gly Phe Gly Lys Leu Pro Glu Lys Ser
165 170 175

Val Leu Glu Lys Met Lys Ala Glu Leu Arg Glu Ala Leu Pro Leu Ala
180 185 190

Glu Tyr Thr Phe Glu Leu Phe Ala Lys Leu Glu Gln Tyr Ser Glu Val
195 200 205

Glu Gly Pro Ile Thr His Leu Ala Val Lys Pro Arg Gly Asp Ala Tyr
210 215 220

Gly Ile Tyr Gly Asp Tyr Ile Lys Ala Ser Asp Gly Glu Glu Phe Pro
225 230 235 240

Ser Glu Lys Tyr Arg Asp Tyr Ile Lys Glu Phe Val Val Glu His Ser
245 250 255

Phe Ala Lys His Ser His Tyr Lys Gly Arg Pro Phe Met Val Gly Ala
260 265 270

Ile Ser Arg Val Ile Asn Asn Ala Asp Leu Leu Tyr Gly Lys Ala Lys
275 280 285

Glu Leu Tyr Glu Ala Asn Lys Asp Leu Leu Lys Gly Thr Asn Pro Phe
290 295 300

Ala Asn Asn Leu Ala Gln Ala Leu Glu Ile Val Tyr Phe Ile Glu Arg
305 310 315 320

Ala Ile Asp Leu Leu Asp Glu Ala Leu Ala Lys Trp Pro Ile Lys Pro
325 330 335

Arg Asp Glu Val Glu Ile Lys Asp Gly Phe Gly Val Ser Thr Thr Glu
340 345 350

Ala Pro Arg Gly Ile Leu Val Tyr Ala Leu Lys Val Glu Asn Gly Arg
355 360 365

Val Ser Tyr Ala Asp Ile Ile Thr Pro Thr Ala Phe Asn Leu Ala Met
370 375 380

Met Glu Glu His Val Arg Met Met Ala Glu Lys His Tyr Asn Asp Asp
385 390 395 400

Pro Glu Arg Leu Lys Ile Leu Ala Glu Met Val Val Arg Ala Tyr Asp
405 410 415

Pro Cys Ile Ser Cys Ser Val His Val Val Arg Leu
420 425

<210> 119
<211> 555
<212> PRT
<213> Escherichia coli

<400> 119

Met Asn Val Asn Ser Ser Ser Asn Arg Gly Glu Ala Ile Leu Ala Ala
1 5 10 15

Leu Lys Thr Gln Phe Pro Gly Ala Val Leu Asp Glu Glu Arg Gln Thr
20 25 30

Pro Glu Gln Val Thr Ile Thr Val Lys Ile Asn Leu Leu Pro Asp Val
35 40 45

Val Gln Tyr Leu Tyr Tyr Gln His Asp Gly Trp Leu Pro Val Leu Phe
50 55 60

Gly Asn Asp Glu Arg Thr Leu Asn Gly His Tyr Ala Val Tyr Tyr Ala
65 70 75 80

Leu Ser Met Glu Gly Ala Glu Lys Cys Trp Ile Val Val Lys Ala Leu
85 90 95

Val Asp Ala Asp Ser Arg Glu Phe Pro Ser Val Thr Pro Arg Val Pro
100 105 110

Ala Ala Val Trp Gly Glu Arg Glu Ile Arg Asp Met Tyr Gly Leu Ile
115 120 125

Pro Val Gly Leu Pro Asp Gln Arg Arg Leu Val Leu Pro Asp Asp Trp
130 135 140

Pro Glu Asp Met His Pro Leu Arg Lys Asp Ala Met Asp Tyr Arg Leu
145 150 155 160

Arg Pro Glu Pro Thr Thr Asp Ser Glu Thr Tyr Pro Phe Ile Asn Glu
165 170 175

Gly Asn Ser Asp Ala Arg Val Ile Pro Val Gly Pro Leu His Ile Thr
180 185 190

Ser Asp Glu Pro Gly His Phe Arg Leu Phe Val Asp Gly Glu Gln Ile
195 200 205

Val Asp Ala Asp Tyr Arg Leu Phe Tyr Val His Arg Gly Met Glu Lys
210 215 220

Leu Ala Glu Thr Arg Met Gly Tyr Asn Glu Val Thr Phe Leu Ser Asp
225 230 235 240

Arg Val Cys Gly Ile Cys Gly Phe Ala His Ser Val Ala Tyr Thr Asn
245 250 255

Ser Val Glu Asn Ala Leu Gly Ile Glu Val Pro Gln Arg Ala His Thr
260 265 270

Ile Arg Ser Ile Leu Leu Glu Val Glu Arg Leu His Ser His Leu Leu
275 280 285

Asn Leu Gly Leu Ser Cys His Phe Val Gly Phe Asp Thr Gly Phe Met
290 295 300

Gln Phe Phe Arg Val Arg Glu Lys Ser Met Thr Met Ala Glu Leu Leu
305 310 315 320

Ile Gly Ser Arg Lys Thr Tyr Gly Leu Asn Leu Ile Gly Gly Val Arg
325 330 335

Arg Asp Ile Leu Lys Glu Gln Arg Leu Gln Thr Leu Lys Leu Val Arg
340 345 350

Glu Met Arg Ala Asp Val Ser Glu Leu Val Glu Met Leu Leu Ala Thr
355 360 365

Pro Asn Met Glu Gln Arg Thr Gln Gly Ile Gly Ile Leu Asp Arg Gln
370 375 380

Ile Ala Arg Asp Leu Arg Phe Asp His Pro Tyr Ala Asp Tyr Gly Asn
385 390 395 400

Ile Pro Lys Thr Leu Phe Thr Phe Thr Gly Gly Asp Val Phe Ser Arg
405 410 415

Val Met Val Arg Val Lys Glu Thr Phe Asp Ser Leu Ala Met Leu Glu
420 425 430

Phe Ala Leu Asp Asn Met Pro Asp Thr Pro Leu Leu Thr Glu Gly Phe
435 440 445

Ser Tyr Lys Pro His Ala Phe Ala Leu Gly Phe Val Glu Ala Pro Arg
450 455 460

Gly Glu Asp Val His Trp Ser Met Leu Gly Asp Asn Gln Lys Leu Phe
465 470 475 480

Arg Trp Arg Cys Arg Ala Ala Thr Tyr Ala Asn Trp Pro Val Leu Arg
485 490 495

Tyr Met Leu Arg Gly Asn Thr Val Ser Asp Ala Pro Leu Ile Ile Gly
500 505 510

Ser Leu Asp Pro Cys Tyr Ser Cys Thr Asp Arg Val Thr Leu Val Asp
515 520 525

Val Arg Lys Arg Gln Ser Lys Thr Val Pro Tyr Lys Glu Ile Glu Arg
530 535 540

Tyr Gly Ile Asp Arg Asn Arg Ser Pro Leu Lys
545 550 555

<210> 120
<211> 405
<212> PRT
<213> Methanothermobacter thermautotrophicus

<400> 120

Met Ser Glu Arg Ile Val Ile Ser Pro Thr Ser Arg Gln Glu Gly His
1 5 10 15

Ala Glu Leu Val Met Glu Val Asp Asp Glu Gly Ile Val Thr Lys Gly
20 25 30

Arg Tyr Phe Ser Ile Thr Pro Val Arg Gly Leu Glu Lys Ile Val Thr
35 40 45

Gly Lys Ala Pro Glu Thr Ala Pro Val Ile Val Gln Arg Ile Cys Gly
50 55 60

Val Cys Pro Ile Pro His Thr Leu Ala Ser Val Glu Ala Ile Asp Asp
65 70 75 80

Ser Leu Asp Ile Glu Val Pro Lys Ala Gly Arg Leu Leu Arg Glu Leu
85 90 95

Thr Leu Ala Ala His His Val Asn Ser His Ala Ile His His Phe Leu
100 105 110

Ile Ala Pro Asp Phe Val Pro Glu Asn Leu Met Ala Asp Ala Ile Asn
115 120 125

Ser Val Ser Glu Ile Arg Lys Asn Ala Gln Tyr Val Val Asp Met Val
130 135 140

Ala Gly Glu Gly Ile His Pro Ser Asp Val Arg Ile Gly Gly Met Ala
145 150 155 160

Asp Asn Ile Thr Glu Leu Ala Arg Lys Arg Leu Tyr Ala Arg Leu Lys
165 170 175

Gln Leu Lys Pro Lys Val Asp Glu His Val Glu Leu Met Ile Gly Leu
180 185 190

Ile Glu Asp Lys Gly Leu Pro Lys Gly Leu Gly Val His Asn Gln Pro
195 200 205

Thr Leu Ala Ser His Gln Ile Tyr Gly Asp Arg Thr Lys Phe Asp Leu
210 215 220

Asp Arg Phe Thr Glu Val Met Pro Glu Ser Trp Tyr Asp Asp Pro Glu
225 230 235 240

Ile Ala Lys Arg Ala Cys Ser Thr Ile Pro Leu Tyr Asp Gly Arg Asn
245 250 255

Val Glu Val Gly Pro Arg Ala Arg Met Val Glu Phe Gln Gly Phe Lys
260 265 270

Glu Arg Gly Val Val Ala Gln His Val Ala Arg Ala Leu Glu Met Lys
275 280 285

Thr Ala Leu Ala Arg Ala Ile Glu Ile Leu Asp Glu Leu Asp Thr Ser
290 295 300

Ala Pro Val Arg Ala Asp Phe Asp Glu Arg Gly Thr Gly Lys Leu Gly
305 310 315 320

Val Gly Ala Ile Glu Gly Pro Arg Gly Leu Asp Val His Met Ala Gln
325 330 335

Val Glu Asn Gly Lys Ile Gln Phe Tyr Ser Ala Leu Val Pro Thr Thr
340 345 350

Trp Asn Ile Pro Thr Met Gly Pro Ala Thr Glu Gly Phe His His Glu
355 360 365

Tyr Gly Pro His Val Ile Arg Ala Tyr Asp Pro Cys Leu Ser Cys Ala
370 375 380

Thr His Val Met Val Val Asp Asp Glu Asp Arg Ser Val Ile Arg Asp
385 390 395 400

Glu Met Val Arg Leu
405

<210> 121
<211> 456
<212> PRT
<213> Methanosarcina barkeri

<400> 121

Met Thr Lys Val Val Glu Ile Ser Pro Thr Thr Arg His Glu Gly His
1 5 10 15

Ser Lys Leu Thr Leu Lys Val Asn Asp Glu Gly Ile Val Glu Arg Gly
20 25 30

Asp Trp Leu Ser Thr Thr Pro Val Arg Gly Ile Glu Lys Leu Ala Ile
35 40 45

Gly Lys Thr Met Asp Gln Val Pro Lys Ile Ala Ser Arg Val Cys Gly
50 55 60

```
Ile Cys Pro Ile Ala His Thr Leu Ala Gly Ile Glu Ala Met Glu Ala
65                  70                  75                  80

Ser Ile Gly Cys Glu Ile Pro Lys Asp Ala Lys Leu Leu Arg Val Ile
                85                  90                  95

Leu His Ala Ala Asn Arg Leu His Ser His Ala Leu His Asn Ile Leu
            100                 105                 110

Ile Leu Pro Asp Phe Tyr Ile Pro Asp Thr Glu Thr Lys Ile Asn Pro
        115                 120                 125

Phe Ser Lys Glu Gln Pro Leu Arg Ser Val Ala Val Arg Ile Phe Arg
    130                 135                 140

Ile Arg Glu Ile Ala Gln Thr Ile Gly Ala Val Ala Gly Gly Glu Ala
145                 150                 155                 160

Ile His Pro Ser Asn Pro Arg Val Gly Gly Met Tyr Arg Asn Val Ser
                165                 170                 175

Ser Arg Ala Lys Gln Lys Ile Ala Asp Leu Ala Lys Glu Gly Leu Val
            180                 185                 190

Leu Ala His Glu Gln Met Glu Phe Met Ile Glu Val Ile Arg Asn Met
        195                 200                 205

Gln Asp Arg Glu Phe Val Glu Val Ala Gly Lys Gln Ile Pro Leu Pro
    210                 215                 220

Lys Thr Leu Gly Tyr His Asn Gln Gly Val Met Ala Thr Ala Pro Met
225                 230                 235                 240

Tyr Gly Ser Ser Ser Leu Asp Glu Lys Pro Met Trp Asp Phe Thr Arg
                245                 250                 255

Trp Arg Glu Thr Arg Pro Trp Asp Trp Tyr Met Ser Glu Glu Thr Ile
            260                 265                 270

Asp Leu Glu Asp Ser Ser Tyr Pro Ile Gly Gly Thr Thr Lys Val Gly
        275                 280                 285

Thr Lys Val Asn Pro Arg Met Glu Ala Cys Asn Thr Val Pro Thr Tyr
    290                 295                 300

Asp Gly Gln Pro Val Glu Val Gly Pro Arg Ala Arg Leu Ala Thr Phe
305                 310                 315                 320

Lys His Phe Thr Glu Lys Gly Thr Phe Ala Gln His Ile Ala Arg Gln
                325                 330                 335

Met Glu Tyr Thr Asp Cys Tyr Tyr Thr Ile Leu Asn Cys Leu Glu Asn
            340                 345                 350

Leu Asp Thr Ser Gly Lys Val Leu Ala Asp Thr Ile Pro Leu Gly Asn
        355                 360                 365
```

Gly Ser Met Gly Trp Ala Ala Asn Glu Ala Pro Arg Gly Thr Asp Val
370 375 380

His Leu Ala Arg Val Lys Asp Gly Lys Val Leu Arg Tyr Glu Met Leu
385 390 395 400

Val Pro Thr Thr Trp Asn Phe Pro Thr Cys Ser Arg Ala Leu Thr Gly
405 410 415

Ala Pro Trp Gln Ile Ala Glu Met Val Ile Arg Ala Tyr Asp Pro Cys
420 425 430

Val Ser Cys Ala Thr His Met Ile Val Val Asn Glu Glu Asp Arg Ile
435 440 445

Val Ala Gln Lys Leu Met Gln Trp
450 455

<210> 122
<211> 361
<212> PRT
<213> Rhodospirillum rubrum

<400> 122

Met Ser Thr Tyr Thr Ile Pro Val Gly Pro Leu His Val Ala Leu Glu
1 5 10 15

Glu Pro Met Tyr Phe Arg Ile Glu Val Asp Gly Glu Lys Val Val Ser
20 25 30

Val Asp Ile Thr Ala Gly His Val His Arg Gly Ile Glu Tyr Leu Ala
35 40 45

Thr Lys Arg Asn Ile Tyr Gln Asn Ile Val Leu Thr Glu Arg Val Cys
50 55 60

Ser Leu Cys Ser Asn Ser His Pro Gln Thr Tyr Cys Met Ala Leu Glu
65 70 75 80

Ser Ile Thr Gly Met Val Val Pro Pro Arg Ala Gln Tyr Leu Arg Val
85 90 95

Ile Ala Asp Glu Thr Lys Arg Val Ala Ser His Met Phe Asn Val Ala
100 105 110

Ile Leu Ala His Ile Val Gly Phe Asp Ser Leu Phe Met His Val Met
115 120 125

Glu Ala Arg Glu Ile Met Gln Asp Thr Lys Glu Ala Val Phe Gly Asn
130 135 140

Arg Met Asp Ile Ala Ala Met Ala Ile Gly Gly Val Lys Tyr Asp Leu
145 150 155 160

Asp Lys Asp Gly Arg Asp Tyr Phe Ile Gly Gln Leu Asp Lys Leu Glu
165 170 175

```
Pro Thr Leu Arg Asp Glu Ile Ile Pro Leu Tyr Gln Thr Asn Pro Ser
            180                 185                 190

Ile Val Asp Arg Thr Arg Gly Ile Gly Val Leu Ser Ala Ala Asp Cys
        195                 200                 205

Val Asp Tyr Gly Leu Met Gly Pro Val Ala Arg Gly Ser Gly His Ala
    210                 215                 220

Tyr Asp Val Arg Lys Gln Ala Pro Tyr Ala Val Tyr Asp Arg Leu Asp
225                 230                 235                 240

Phe Glu Met Ala Leu Gly Glu His Gly Asp Val Trp Ser Arg Ala Met
                245                 250                 255

Val Arg Trp Gln Glu Ala Leu Thr Ser Ile Gly Leu Ile Arg Gln Cys
            260                 265                 270

Leu Arg Asp Met Pro Asp Gly Pro Thr Lys Ala Gly Pro Val Pro Pro
        275                 280                 285

Ile Pro Ala Gly Glu Ala Val Ala Lys Thr Glu Ala Pro Arg Gly Glu
    290                 295                 300

Leu Ile Tyr Tyr Leu Lys Thr Asn Gly Thr Asp Arg Pro Glu Arg Leu
305                 310                 315                 320

Lys Trp Arg Val Pro Thr Tyr Met Asn Trp Asp Ala Leu Asn Val Met
                325                 330                 335

Met Ala Gly Ala Arg Ile Ser Asp Ile Pro Leu Ile Val Asn Ser Ile
            340                 345                 350

Asp Pro Cys Ile Ser Cys Thr Glu Arg
        355                 360


<210>  123
<211>  505
<212>  PRT
<213>  Artificial sequence

<220>
<223>  Synthetic sequence

<400>  123

Met Ala Leu Gly Leu Leu Ala Glu Leu Arg Ala Gly Gln Ala Val Ala
1               5                   10                  15

Cys Ala Arg Arg Thr Asn Ala Pro Ala His Pro Ala Ala Val Val Pro
            20                  25                  30

Cys Leu Pro Ser Arg Ala Gly Lys Phe Phe Asn Leu Ser Gln Lys Val
        35                  40                  45

Pro Ser Ser Gln Ser Ala Arg Gly Ser Thr Ile Arg Val Ala Ala Thr
    50                  55                  60
```

```
Ala Thr Asp Ala Val Pro His Trp Lys Leu Ala Leu Glu Glu Leu Asp
65                  70                  75                  80

Lys Pro Lys Asp Gly Gly Arg Lys Val Leu Ile Ala Gln Val Ala Pro
                85                  90                  95

Ala Val Arg Val Ala Ile Ala Glu Ser Phe Gly Leu Ala Pro Gly Ala
            100                 105                 110

Val Ser Pro Gly Lys Leu Ala Thr Gly Leu Arg Ala Leu Gly Phe Asp
        115                 120                 125

Gln Val Phe Asp Thr Leu Phe Ala Ala Asp Leu Thr Ile Trp Glu Glu
    130                 135                 140

Gly Thr Glu Leu Leu His Arg Leu Lys Glu His Leu Glu Ala His Pro
145                 150                 155                 160

His Ser Asp Glu Pro Leu Pro Met Phe Thr Ser Cys Cys Pro Gly Trp
                165                 170                 175

Val Ala Met Met Glu Lys Ser Tyr Pro Glu Leu Ile Pro Phe Val Ser
            180                 185                 190

Ser Cys Lys Ser Pro Gln Met Met Met Gly Ala Met Val Lys Thr Tyr
        195                 200                 205

Leu Ser Glu Lys Gln Gly Ile Pro Ala Lys Asp Ile Val Met Val Ser
    210                 215                 220

Val Met Pro Cys Val Arg Lys Gln Gly Glu Ala Asp Arg Glu Trp Phe
225                 230                 235                 240

Cys Val Ser Glu Pro Gly Val Arg Asp Val Asp His Val Ile Thr Thr
                245                 250                 255

Ala Glu Leu Gly Asn Ile Phe Lys Glu Arg Gly Ile Asn Leu Pro Glu
            260                 265                 270

Leu Pro Asp Ser Asp Trp Asp Gln Pro Leu Gly Leu Gly Ser Gly Ala
        275                 280                 285

Gly Val Leu Phe Gly Thr Thr Gly Gly Val Met Glu Ala Ala Leu Arg
    290                 295                 300

Thr Ala Tyr Glu Ile Val Thr Lys Glu Pro Leu Pro Arg Leu Asn Leu
305                 310                 315                 320

Ser Glu Val Arg Gly Leu Asp Gly Ile Lys Glu Ala Ser Val Thr Leu
                325                 330                 335

Val Pro Ala Pro Gly Ser Lys Phe Ala Glu Leu Val Ala Glu Arg Leu
            340                 345                 350

Ala His Lys Val Glu Glu Ala Ala Ala Ala Glu Ala Ala Ala Ala Val
        355                 360                 365
```

Glu Gly Ala Val Lys Pro Pro Ile Ala Tyr Asp Gly Gly Gln Gly Phe
370 375 380

Ser Thr Asp Asp Gly Lys Gly Gly Leu Lys Leu Arg Val Ala Val Ala
385 390 395 400

Asn Gly Leu Gly Asn Ala Lys Lys Leu Ile Gly Lys Met Val Ser Gly
405 410 415

Glu Ala Lys Tyr Asp Phe Val Glu Ile Met Ala Cys Pro Ala Gly Cys
420 425 430

Val Gly Gly Gly Gly Gln Pro Arg Ser Thr Asp Lys Gln Ile Thr Gln
435 440 445

Lys Arg Gln Ala Ala Leu Tyr Asp Leu Asp Glu Arg Asn Thr Leu Arg
450 455 460

Arg Ser His Glu Asn Glu Ala Val Asn Gln Leu Tyr Lys Glu Phe Leu
465 470 475 480

Gly Glu Pro Leu Ser His Arg Ala His Glu Leu Leu His Thr His Tyr
485 490 495

Val Pro Gly Gly Ala Glu Ala Asp Ala
500 505

<210> 124
<211> 19
<212> PRT
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 124

Gly Ala Gly Val Ile Phe Gly Ala Thr Gly Gly Val Met Glu Ala Ala
1 5 10 15

Leu Arg Thr

<210> 125
<211> 19
<212> PRT
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 125

Gly Gly Gly Ala Ile Phe Cys Ala Thr Gly Gly Val Met Glu Ala Ala
1 5 10 15

Val Arg Ser

<210> 126

<211> 19
<212> PRT
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 126

Gly Gly Ala Thr Ile Phe Gly Val Thr Gly Gly Val Met Glu Ala Ala
1 5 10 15

Leu Arg Phe

<210> 127
<211> 19
<212> PRT
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 127

Gly Ala Gly Ala Ile Phe Gly Ala Thr Gly Gly Val Met Glu Ala Ala
1 5 10 15

Leu Arg Ser

<210> 128
<211> 19
<212> PRT
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 128

Gly Ala Gly Ala Ile Phe Gly Ala Thr Gly Gly Val Met Glu Ala Ala
1 5 10 15

Ile Arg Ser

<210> 129
<211> 19
<212> PRT
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 129

Gly Ala Ala Val Ile Phe Gly Val Thr Gly Gly Val Met Glu Ala Ala
1 5 10 15

Leu Arg Thr

<210> 130
<211> 19
<212> PRT
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 130

Gly Ala Gly Gln Ile Phe Ala Ala Thr Gly Gly Val Met Glu Ala Ala
1 5 10 15

Ser Arg Thr

<210> 131
<211> 19
<212> PRT
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 131

Gly Ala Ala Val Ile Phe Gly Thr Thr Gly Gly Val Met Glu Ala Ala
1 5 10 15

Leu Arg Thr

<210> 132
<211> 19
<212> PRT
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 132

Gly Ala Ala Pro Ile Phe Gly Val Thr Gly Gly Val Ile Glu Ala Ala
1 5 10 15

Leu Arg Thr

<210> 133
<211> 19
<212> PRT
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 133

Gly Ala Gly Val Ile Phe Gly Thr Thr Gly Gly Val Met Glu Ala Ala
1 5 10 15

Leu Arg Ser

<210> 134
<211> 19
<212> PRT
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 134

Gly Ala Gly Val Ile Phe Gly Ala Thr Gly Gly Val Met Glu Ala Ala
1 5 10 15

Ile Arg Thr

<210> 135
<211> 19
<212> PRT
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 135

Ser Ala Gly Asn Leu Phe Gly Val Thr Gly Gly Val Met Glu Ala Ala
1 5 10 15

Ile Arg Thr

<210> 136
<211> 19
<212> PRT
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 136

Gly Ala Gly Ala Ile Phe Gly Ala Thr Gly Gly Val Met Glu Ala Ala
1 5 10 15

Leu Arg Thr

<210> 137
<211> 19
<212> PRT
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 137

Gly Ala Gly Val Leu Phe Gly Thr Thr Gly Gly Val Met Glu Ala Ala
1 5 10 15

Leu Arg Thr

<210> 138
<211> 19
<212> PRT
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 138

Gly Ala Ala Ala Leu Phe Gly Val Thr Gly Gly Val Met Glu Ala Ala
1 5 10 15

Leu Arg Thr

<210> 139
<211> 19
<212> PRT
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 139

Gly Ala Gly Val Leu Phe Gly Thr Thr Gly Gly Val Met Glu Ala Ala
1 5 10 15

Val Arg Thr

<210> 140
<211> 19
<212> PRT
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 140

Gly Ala Gly Thr Ile Phe Gly Thr Thr Gly Gly Val Met Glu Ala Ala
1 5 10 15

Leu Arg Thr

<210> 141
<211> 19
<212> PRT
<213> Artificial sequence

<220>
<223> Synthetic construct

<400> 141

Gly Gly Gly Val Leu Phe Gly Thr Thr Gly Gly Val Met Glu Ala Ala
1 5 10 15

Leu Arg Thr

<210> 142
<211> 5
<212> PRT
<213> Artificial sequence

<220>
<223> Synthetic construct

<400> 142

Thr Ile Met Glu Glu
1 5

<210> 143
<211> 5
<212> PRT
<213> Artificial sequence

<220>
<223> Synthetic construct

<400> 143

Thr Ile Val Glu Glu
1 5

<210> 144
<211> 5
<212> PRT
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 144

Thr Ile Trp Glu Glu
1 5

<210> 145
<211> 5
<212> PRT
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 145

Thr Ile Cys Glu Glu
1 5

<210> 146
<211> 5
<212> PRT
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 146

Val Ile Met Glu Glu
1 5

<210> 147
<211> 5
<212> PRT
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 147

Thr Ala Arg Leu Glu
1 5

<210> 148
<211> 260
<212> DNA
<213> Chlamydomonas reinhardtii

<400> 148
gcagttgggt caggggctgg cgacgcgctg ctgacgcgca agtgaatggc ccaacaagtc 60

gcctcgcggt cgctgtcggc gccaaacccg cagctgcatc caccagattc acttgttaga 120

tcgacctagg ttgcgggacc ggaggcggct cgctgtgcaa gcgcggtgac ctcgtacggc 180

ggcatggatc gccatctcga ttcgcgcggc agaatcgggc cccgcgcaca tttaagccgc 240

gggcgagact catttcgtta 260

<210> 149
<211> 1181
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 149
gccagaagga gcgcagccaa accaggatga tgtttgatgg ggtatttgag cacttgcaac 60

ccttatccgg aagccccctg gcccacaaag gctaggcgcc aatgcaagca gttcgcatgc 120

agcccctgga gcggtgccct cctgataaac cggccagggg gcctatgttc tttacttttt 180

tacaagagaa gtcactcaac atcttaaaat ggccaggtga gtcgacgagc aagcccggcg 240

gatcaggcag cgtgcttgca gatttgactt gcaacgcccg cattgtgtcg acgaaggctt 300

ttggctcctc tgtcgctgtc tcaagcagca tctaaccctg cgtcgccgtt tccatttgca 360

ggatggccaa gctgaccagc gccgttccgg tgctcaccgc gcgcgacgtc gccggagcgg 420

tcgagttctg gaccgaccgg ctcgggttct cccgggactt cgtggaggac gacttcgccg 480

gtgtggtccg ggacgacgtg accctgttca tcagcgcggt ccaggaccag gtgagtcgac 540

gagcaagccc ggcggatcag gcagcgtgct tgcagatttg acttgcaacg cccgcattgt 600

gtcgacgaag gcttttggct cctctgtcgc tgtctcaagc agcatctaac cctgcgtcgc 660

cgtttccatt tgcaggacca ggtggtgccg gacaacaccc tggcctgggt gtgggtgcgc 720

ggcctggacg agctgtacgc cgagtggtcg gaggtcgtgt ccacgaactt ccgggacgcc 780

tccgggccgg ccatgaccga gatcggcgag cagccgtggg ggcgggagtt cgccctgcgc 840

gacccggccg gcaactgcgt gcacttcgtg gccgaggagc aggactaacc gacgtcgacc 900

cactctagag gatcgatccc cgctccgtgt aaatggaggc gctcgttgat ctgagccttg 960

ccccctgacg aacggcggtg gatggaagat actgctctca agtgctgaag cggtagctta 1020

gctccccgtt tcgtgctgat cagtcttttt caacacgtaa aaagcggagg agttttgcaa 1080

ttttgttggt tgtaacgatc ctccgttgat tttggcctct ttctccatgg gcgggctggg 1140

cgtatttgaa gcttaattaa ctcgaggggg ggcccggtac c 1181

<210> 150
<211> 260
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 150
ccgacgtcga cccactctag aggatcgatc cccgctccgt gtaaatggag gcgctcgttg 60

atctgagcct tgccccctga cgaacggcgg tggatggaag atactgctct caagtgctga 120

agcggtagct tagctccccg tttcgtgctg atcagtcttt ttcaacacgt aaaaagcgga 180

ggagttttgc aattttgttg gttgtaacga tcctccgttg attttggcct ctttctccat 240

gggcgggctg ggcgtatttg 260

<210> 151
<211> 520
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 151
ccgacgtcga cccactctag aggatcgatc cccgctccgt gtaaatggag gcgctcgttg 60

atctgagcct tgccccctga cgaacggcgg tggatggaag atactgctct caagtgctga 120

agcggtagct tagctccccg tttcgtgctg atcagtcttt ttcaacacgt aaaaagcgga 180

ggagttttgc aattttgttg gttgtaacga tcctccgttg attttggcct ctttctccat 240

gggcgggctg ggcgtatttg gcagttgggt caggggctgg cgacgcgctg ctgacgcgca 300

agtgaatggc ccaacaagtc gcctcgcggt cgctgtcggc gccaaacccg cagctgcatc 360

caccagattc acttgttaga tcgacctagg ttgcgggacc ggaggcggct cgctgtgcaa 420

gcgcggtgac ctcgtacggc ggcatggatc gccatctcga ttcgcgcggc agaatcgggc 480

cccgcgcaca tttaagccgc gggcgagact catttcgtta 520

<210> 152
<211> 30
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 152
atccgtagtt atccttatgg ccatcttagc 30

<210> 153
<211> 30
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 153
cgtgcatcga ttaacagctt ctggacctga 30

<210> 154
<211> 30
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 154
ttaaacgtcg tacgtccaag tataactaag 30

<210> 155
<211> 30
<212> DNA

<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 155
aatctgatac atgctattca gatcttacaa 30

<210> 156
<211> 30
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 156
tcttccatcg taaatctagc atcgattagc 30

<210> 157
<211> 30
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 157
atctgtaata atctagtcga ggcattcaag 30

<210> 158
<211> 30
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 158
aactggctta aatcgttaac aatcgtgtga 30

<210> 159
<211> 30
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 159
gatttaacat aactgtcgat taccgtgcga 30

<210> 160
<211> 30
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 160
tatgcttgac aatcgtaatc ctggtgacaa 30

<210> 161
<211> 30
<212> DNA
<213> Artificial sequence

<220>

<223> Synthetic sequence

<400> 161
taacaagaat ctggctaatc aatcgatgca 30

<210> 162
<211> 30
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 162
gtagtcggaa tagttactaa cgaggattcg 30

<210> 163
<211> 30
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 163
aaatgtctac tcgactagta aatcgtaact 30

<210> 164
<211> 290
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 164
gcagttgggt caggggctgg cgacgcgctg ctgacgcgca agtgaatggc ccaacaagtc 60

gcctcgcggt cgctgtcggc gccaaacccg cagctgcatc caccagattc acttgttaga 120

tcgacctagg ttgcgggacc ggaggcggct cgctgtgcaa gcgcggtgac ctcgtacggc 180

ggcatggatc gccatctcga ttcgcgcggc agaatcgggc cccgcgcaca tttaagccgc 240

gggcgagact catttcgtta atccgtagtt atccttatgg ccatcttagc 290

<210> 165
<211> 580
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 165
cgtgcatcga ttaacagctt ctggacctga ccgacgtcga cccactctag aggatcgatc 60

cccgctccgt gtaaatggag gcgctcgttg atctgagcct tgccccctga cgaacggcgg 120

tggatggaag atactgctct caagtgctga agcggtagct tagctccccg tttcgtgctg 180

atcagtcttt ttcaacacgt aaaaagcgga ggagttttgc aattttgttg gttgtaacga 240

tcctccgttg attttggcct ctttctccat gggcgggctg ggcgtatttg gcagttgggt 300

caggggctgg cgacgcgctg ctgacgcgca agtgaatggc ccaacaagtc gcctcgcggt 360

cgctgtcggc gccaaacccg cagctgcatc caccagattc acttgttaga tcgacctagg 420

ttgcgggacc ggaggcggct cgctgtgcaa gcgcggtgac ctcgtacggc ggcatggatc 480

gccatctcga ttcgcgcggc agaatcgggc cccgcgcaca tttaagccgc gggcgagact 540

catttcgtta ttaaacgtcg tacgtccaag tatgactaag 580

<210> 166
<211> 566
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 166
aatctgatac atgctattca gatcttacaa ccgacgtcga cccactctag aggatcgatc 60

cccgctccgt gtaaatggag gcgctcgttg atctgagcct tgccccctga cgaacggcgg 120

tggatggaag atactgctct caagtgctga agcggtagct tagctccccg tttcgtgctg 180

atcagtcttt ttcaacacgt aaaaagcgga ggagttttgc aattttgttg gttgtaacga 240

tcctccgttg attttggcct ctttctccat gggcgggctg ggcgtatttg gcagttgggt 300

caggggctgg cgacgcgctg ctgacgcgca agtgaatggc ccaacaagtc gcctcgcggt 360

cgctgtcggc gccaaacccg cagctgcatc caccagattc acttgttaga tcgacctagg 420

ttgcgggacc ggaggcggct cgctgtgcaa gcgcggtgac ctcgtacggc ggcatggatc 480

gccatctcga ttcgcgcggc agaatcgggc cccgcgcaca tttaagccgc gggcgatctt 540

ccatcgtaaa tctagcatcg attagc 566

<210> 167
<211> 290
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 167
atctgtaata atctagtcga ggcattcaag ccgacgtcga cccactctag aggatcgatc 60

cccgctccgt gtaaatggag gcgctcgttg atctgagcct tgccccctga cgaacggcgg 120

tggatggaag atactgctct caagtgctga agcggtagct tagctccccg tttcgtgctg 180

atcagtcttt ttcaacacgt aaaaagcgga ggagttttgc aattttgttg gttgtaacga 240

tcctccgttg attttggcct ctttctccat gggcgggctg ggcgtatttg 290

<210> 168
<211> 1181
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 168
gccagaagga gcgcagccaa accaggatga tgtttgatgg ggtatttgag cacttgcaac 60

ccttatccgg aagccccctg gcccacaaag gctaggcgcc aatgcaagca gttcgcatgc 120

agcccctgga gcggtgccct cctgataaac cggccagggg gcctatgttc tttacttttt 180

tacaagagaa gtcactcaac atcttaaaat ggccaggtga gtcgacgagc aagcccggcg 240

gatcaggcag cgtgcttgca gatttgactt gcaacgcccg cattgtgtcg acgaaggctt 300

ttggctcctc tgtcgctgtc tcaagcagca tctaaccctg cgtcgccgtt tccatttgca 360

ggatggccaa gctgaccagc gccgttccgg tgctcaccgc gcgcgacgtc gccggagcgg 420
tcgagttctg gaccgaccgg ctcgggttct cccgggactt cgtggaggac gacttcgccg 480
gtgtggtccg ggacgacgtg accctgttca tcagcgcggt ccaggaccag gtgagtcgac 540
gagcaagccc ggcggatcag gcagcgtgct tgcagatttg acttgcaacg cccgcattgt 600
gtcgacgaag gcttttggct cctctgtcgc tgtctcaagc agcatctaac cctgcgtcgc 660
cgtttccatt tgcaggacca ggtggtgccg gacaacaccc tggcctgggt gtgggtgcgc 720
ggcctggacg agctgtacgc cgagtggtcg gaggtcgtgt ccacgaactt ccgggacgcc 780
tccgggccgg ccatgaccga gatcggcgag cagccgtggg ggcgggagtt cgccctgcgc 840
gacccggccg gcaactgcgt gcacttcgtg gccgaggagc aggactaacc gacgtcgacc 900
cactctagag gatcgatccc cgctccgtgt aaatggaggc gctcgttgat ctgagccttg 960
ccccctgacg aacggcggtg gatggaagat actgctctca agtgctgaag cggtagctta 1020
gctccccgtt tcgtgctgat cagtcttttt caacacgtaa aaagcggagg agttttgcaa 1080
ttttgttggt tgtaacgatc ctccgttgat tttggcctct ttctccatgg gcgggctggg 1140
cgtatttgaa gcttaattaa ctcgaggggg ggcccggtac c 1181

<210> 169
<211> 290
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 169
gcagttgggt caggggctgg cgacgcgctg ctgacgcgca agtgaatggc ccaacaagtc 60
gcctcgcggt cgctgtcggc gccaaacccg cagctgcatc caccagattc acttgttaga 120
tcgacctagg ttgcgggacc ggaggcggct cgctgtgcaa gcgcggtgac ctcgtacggc 180
ggcatggatc gccatctcga ttcgcgcggc agaatcgggc cccgcgcaca tttaagccgc 240
gggcgagact catttcgtta aactggctta aatcgttaac aatcgtgtga 290

<210> 170
<211> 566
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 170
gatttaacat aactgtcgat taccgtgcga ccgacgtcga cccactctag aggatcgatc 60
cccgctccgt gtaaatggag gcgctcgttg atctgagcct tgccccctga cgaacggcgg 120
tggatggaag atactgctct caagtgctga agcggtagct tagctccccg tttcgtgctg 180
atcagtcttt ttcaacacgt aaaaagcgga ggagttttgc aattttgttg gttgtaacga 240
tcctccgttg attttggcct ctttctccat gggcgggctg ggcgtatttg gcagttgggt 300
caggggctgg cgacgcgctg ctgacgcgca agtgaatggc ccaacaagtc gcctcgcggt 360
cgctgtcggc gccaaacccg cagctgcatc caccagattc acttgttaga tcgacctagg 420
ttgcgggacc ggaggcggct cgctgtgcaa gcgcggtgac ctcgtacggc ggcatggatc 480

gccatctcga ttcggcgcgg cagaatcggg cccgcgcaca tttaagccgc gggcgatatg 540

cttgacaatc gtaatcctgg tgacaa 566

<210> 171
<211> 290
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 171
taacaagaat ctggctaatc aatcgatgca ccgacgtcga cccactctag aggatcgatc 60

cccgctccgt gtaaatggag gcgctcgttg atctgagcct tgccccctga cgaacggcgg 120

tggatggaag atactgctct caagtgctga agcggtagct tagctccccg tttcgtgctg 180

atcagtcttt ttcaacacgt aaaaagcgga ggagttttgc aattttgttg gttgtaacga 240

tcctccgttg attttggcct ctttctccat gggcgggctg ggcgtatttg 290

<210> 172
<211> 381
<212> DNA
<213> Chlamydomonas reinhardtii

<400> 172
atggccatgg ctatgcgctc caccttcgcc gcccgcgttg gcgctaagcc cgctgtccgc 60

ggtgctcgcc ccgccagccg catgagctgc atggcctaca aggtcaccct gaagacccct 120

tcgggcgaca agaccattga gtgccccgct gacacctaca tcctggacgc tgctgaggag 180

gccggcctgg acctgcccta ctcttgccgc gctggtgctt gctccagctg cgccggcaag 240

gtcgctgccg gcaccgtcga ccagtcggac cagtccttcc tggacgatgc ccagatgggc 300

aacggcttcg tgctgacctg cgtggcctac cccacctcgg actgcaccat ccagacccac 360

caggaggagg ccctgtacta a 381

<210> 173
<211> 1494
<212> DNA
<213> Chlamydomonas reinhardtii

<400> 173
atgtcggcgc tcgtgctgaa gccctgcgcg gccgtgtcta ttcgcggcag ctcctgcagg 60

gcgcggcagg tcgcccccg cgctccgctc gcagccagca ccgtgcgtgt agcccttgca 120

acacttgagg cgcccgcacg ccgcctaggc aacgtcgctt gcgcggctgc cgcacccgct 180

gcggaggcgc ctttgagtca tgtccagcag gcgctcgccg agcttgccaa gcccaaggac 240

gaccccacgc gcaagcacgt ctgcgtgcag gtggctccgg ccgttcgtgt cgctattgcc 300

gagaccctgg gcctggcgcc gggcgccacc acccccaagc agctggccga gggcctccgc 360

cgcctcggct ttgacgaggt gtttgacacg ctgtttggcg ccgacctgac catcatggag 420

gagggcagcg agctgctgca ccgcctcacc gagcacctgg aggcccaccc gcactccgac 480

gagccgctgc ccatgttcac cagctgctgc cccggctgga tcgctatgct ggagaaatct 540

tacccggacc tgatccccta cgtgagcagc tgcaagagcc cccagatgat gctggcggcc 600

atggtcaagt cctacctagc ggaaaagaag ggcatcgcgc caaaggacat ggtcatggtg 660

tccatcatgc cctgcacgcg caagcagtcg gaggctgacc gcgactggtt ctgtgtggac 720

gccgacccca ccctgcgcca gctggaccac gtcatcacca ccgtggagct gggcaacatc 780
ttcaaggagc gcggcatcaa cctggccgag ctgcccgagg gcgagtggga caatccaatg 840
ggcgtgggct cgggcgccgg cgtgctgttc ggcaccaccg gcggtgtcat ggaggcggcg 900
ctgcgcacgg cctatgagct gttcacgggc acgccgctgc cgcgcctgag cctgagcgag 960
gtgcgcggca tggacggcat caaggagacc aacatcacca tggtgcccgc gcccgggtcc 1020
aagtttgagg agctgctgaa gcaccgcgcc gccgcgcgcg ccgaggccgc cgcgcacggc 1080
acccccgggc cgctggcctg ggacggcggc gcgggcttca ccagcgagga cggcaggggc 1140
ggcatcacac tgcgcgtggc cgtggccaac gggctgggca acgccaagaa gctgatcacc 1200
aagatgcagg ccggcgaggc caagtacgac tttgtggaga tcatggcctg ccccgcgggc 1260
tgtgtgggcg gcggcggcca gccccgctcc accgacaagg ccatcacgca gaagcggcag 1320
gcggcgctgt acaacctgga cgagaagtcc acgctgcgcc gcagccacga gaacccgtcc 1380
atccgcgagc tgtacgacac gtacctcgga gagccgctgg gccacaaggc gcacgagctg 1440
ctgcacaccc actacgtggc cggcggcgtg gaggagaagg acgagaagaa gtga 1494

<210> 174
<211> 1725
<212> DNA
<213> Clostriduim pasteuranum

<400> 174
atgaaaacaa taattataaa tggtgtacag tttaatactg atgaagacac tactatatta 60
aaatttgcac gagacaacaa tattgatata tctgcactgt gtttttaaa taattgtaat 120
aatgacataa ataagtgtga aatatgtact gtagaggtag agggtactgg attagtaaca 180
gcctgtgata cattaattga ggatggtatg attataaaca caaattccga tgctgtcaac 240
gaaaaaatta aatctagaat atctcaatta ttagacatac atgaattcaa atgtggtcct 300
tgcaatagaa gagaaaactg tgaattctta aaacttgtta taaaatataa agcaagagct 360
tctaaaccat ttttacctaa agataagact gaatatgtag atgaaagaag taaatcatta 420
actgtagata ggacaaaatg cttattatgt ggaagatgtg ttaatgcctg tggaaaaaat 480
actgaaacct atgcaatgaa atttttaaac aaaaatggta aaactataat tggagcagag 540
gatgaaaaat gctttgatga tactaattgt ctattatgtg gtcaatgtat aatcgcctgt 600
ccagtagcag cattatcgga aaaatcacac atggatagag taaaaaatgc cttaaatgcc 660
cctgaaaaac atgtaatagt agctatggct ccatctgtca gagcttctat aggtgaactt 720
tttaatatgg gatttggcgt tgacgtaaca ggaaaaattt atactgcttt aagacagctt 780
ggatttgata aaatattcga tataaacttc ggagcagata tgacaattat ggaagaggct 840
acagaattag ttcaaagaat agagaataat ggacctttcc caatgtttac atcttgctgc 900
ccaggttggg taagacaagc tgaaaattat tatcctgaat tactaaataa tctttcatca 960
gctaaatcac ctcaacaaat ttttggtact gctagtaaaa cttattatcc ttctatatct 1020
ggtcttgacc caaagaatgt atttactgta acagttatgc cctgtacttc aaaaaaattt 1080
gaagcagata gaccacaaat ggaaaaagac ggcctaagag atatagatgc tgttataact 1140
actcgagaat tagcaaaaat gattaaagat gctaaaatac catttgctaa acttgaagat 1200
agcgaagcag accctgctat gggagaatac agcggtgctg gtgccatatt tggtgcaact 1260

ggcggagtta tggaagcagc tttaagaagt gcaaaagact ttgctgaaaa cgctgaactt 1320
gaagatatag aatataagca agttagagga ttaaatggta taaaagaagc tgaagtagaa 1380
ataaataaca acaaatataa tgtagctgtt ataaatggtg cttcaaattt atttaagttt 1440
atgaaatctg gtatgattaa cgaaaaacaa tatcatttca tagaagtaat ggcttgtcat 1500
ggaggatgtg taaatggtgg tggacagcct catgtaaacc caaaagattt agaaaaagta 1560
gacataaaaa aagtaagagc ttctgtattg tataatcagg atgaacatct ttccaagaga 1620
aaatctcatg aaaatactgc attagttaaa atgtatcaaa attattttgg caaaccaggt 1680
gaaggtcgtg cccatgaaat attacacttt aaatataaaa aataa 1725

<210> 175
<211> 1265
<212> DNA
<213> Desulfovibrio vulgaris

<400> 175
atgagccgta ccgtcatgga gcgcatcgaa tatgagatgc acactccgga ccccaaggcc 60
gatccggaca agctccactt cgtccagatc gacgaggcaa agtgcatagg ctgcgacacc 120
tgttcgcagt actgccccac cgccgccatc ttcggcgaaa tgggcgaacc gcactccatt 180
ccccacatcg aggcgtgcat caactgcggc cagtgcctca cgcactgccc cgagaacgcc 240
atctacgagg cacagtcgtg gtgcctgaag tcgagaagaa gctgaaggac ggcaaggtga 300
aatgcatcgc catgcccgcc cccgccgtgc gctatgcact gggcgacgcc ttcggcatgc 360
ccgtcggttc cgtcaccacc ggcaagatgc tcgcggccct gcagaagctc ggcttcgctc 420
attgctggga caccgagttc accgctgacg tgaccatctg ggaagagggg tccgagttcg 480
tggaacgcct caccaagaag agcgacatgc cgctgccgca gttcacctcg tgctgccccg 540
gctggcagaa gtatgccgag acctactacc ccgaactgct gccgcacttc tccacgtgca 600
agtcgcccat cggcatgaac ggcgcactgg cgaagaccta cggcgcagag cggatgaagt 660
acgaccccaa gcaggtctac accgtctcca tcatgccctg catcgcaaag aagtacgaag 720
ggttgcgtcc cgaactgaag tccagcggca tgcgcgacat cgacgccacg ctgaccaccc 780
gtgagctggc ctacatgatc aagaaggccg gtatcgactt cgcgaaactc cccgacggca 840
agcgtgacag cctcatgggt gaatccaccg gcggtgccac catcttcggc gtcaccggcg 900
gcgtcatgga agcggcactc cgcttcgcct acgaagccgt caccggcaag aagcccgaca 960
gctgggactt caaggccgtg cgcggtcttg atggcatcaa ggaagccacc gtcaacgtcg 1020
gcggtaccga cgtcaaggtc gccgtggtgc acggggccaa gcggttcaag caggtctgcg 1080
acgatgtgaa ggcgggcaag tcgccctatc acttcatcga atacatggcc tgccccggcg 1140
gctgcgtctg tggcggcggt cagcccgtca tgcccggcgt gctcgaagcc atggaccgca 1200
ccaccacccg cctttacgcg ggcctgaaga agcgcctcgc catggcgagc gccaacaagg 1260
catag 1265

<210> 176
<211> 1407
<212> DNA
<213> Entamoeba histolytica

<400> 176

050118 CIP Sequence Listing

atgccaccta aaccatcaca tacactcacc ggacatgacc ataaccatag tattcaattt 60

gattggtcta aatgcatggg ttgtggaatg tgtgctacta aatgtacttt tggggtgtta 120

gtaaaacaac caccaaaaat tccaccattt gttcagccta atagagaaaa actctctcaa 180

gaaaataccg acaagacaag agtacttatt gatgagtctg aatgtactgg gtgtggtcaa 240

tgttctttgg tttgtaactt tggttctatt acaccaatag accatcttgt tgatactttt 300

aaagctaaag aagctggaaa gaagcttgtt gctatgattg caccttcaac tcgtttaggt 360

gttgctgagg ctatgggaat gcctattgga agtacagcta tggctcagtt agttcattgt 420

ttaagactta ttggatttga ttatgtattt gatgttgatg ctggagctga taagacaaca 480

atggatgatt atgccgaagt tattgaaatg aaaaaagaag gaaaaggacc tgctattact 540

tcctgttgtc ctgcttggat tgaacttgtt gaaaaagaat atcctgactt aattccaaac 600

gtctctactg cccgttcacc aattggatgt ttagctggtt gtattaaaag aggatgggca 660

aaggatgtag gaattgcagt agaagatctt tacactgttg gaataatgcc ttgtattgct 720

aaaaaaacag agtctcaaag acaacaaatt catcaagact atgatgcttc atgtacttca 780

aatgaaattg ctgcttattt caaaaaacat cttccacctg aagaatgtaa atttacacaa 840

gaaagagaag aagcacttgc taaaactgaa gatggtcaat gtgatttacc atttagacgt 900

atttctggtg gttctaatat ttttggaaag actggaggag tttgtgaaac tgtattgaga 960

gtaattgcac gtaatgcagg agttgattgg aacagttgta ctgttaacaa ggaagaaact 1020

tttaaacatg ctgcaagtgg atcaacaatg acaaatcttt ctgttgatat tggtggaact 1080

attatcacag gtgctgtttg tcatggtggt tatgctatta gacatgcttg tgaacttatt 1140

agaaaaggag agttaaaagt tgatgttgtt gaaatgatgg catgtgttgg aggttgtctt 1200

ggaggagcag gtcaaccaaa aattccacca gcaaagaaac ttgagatgga taagagaaga 1260

gtaatgttag atattttaga tcaacaaact gatattagag ctgctaatga aaatactgat 1320

gttcttggat ggattgataa acattttgat catcaaggtg cacatcagca tcttcacaca 1380

tattttactc ccagatatca aaactaa 1407

<210> 177
<211> 1350
<212> DNA
<213> Scenedesmus obliquus

<400> 177
atgcctgagt ggcaaccggg aggtcggtat gctgtttctg tccgcccgcc agtgaacagg 60

cgggctgtgg tggcagcaga gcgcaggcgc cttgttgtgc gggcagctgg cccaacagca 120

gaatgtgatt gcccaccagc tcccgcgccc aaggccccgc actggcagca gacgctagat 180

gagctagcca agcctaagga gcagcgcaag gtgatgatcg cccagatcgc accagcagtg 240

cgcgtggcta ttgcagagac catgggactc aaccctgggg atgtgacagt tggccagatg 300

gtgaccggcc tgcgcatgct gggctttgat tatgtgtttg acacgctgtt tggtgctgac 360

ctcaccatca tggaggaggg cacagagcta cggcacaggc ttcaggacca cctggagcag 420

caccccaaca aggaggagcc gctgcccatg ttcaccagct gctgccctgg ctgggtggcc 480

atggtggaga agtccaaccc cgagctcatc ccctacctgt cttcctgcaa gtcgccccag 540

atgatgctgg gcgcagtcat caagaactac ttcgctgccg aggccggcgc caagcctgag 600

gacatctgca acgtgagcgt gatgcctctgc gtgcgcaagc agggcgaggc tgaccgcgag 660
tggttcaaca ccacaggggc tggcggcgcg aacgtggacc acgtcatgac aactgcagag 720
ctgggcaaga tctttgtgga gcgcggaatc aagctgaacg acctgcagga gtcgcccttt 780
gacaaccccg tcggcgaggg cagcggcggc ggcgtgctgt tcggcaccac tggaggcgtg 840
atggaggcgg cgctgcgcac cgtgtacgaa gtggtcacac agaagccttt ggaccgcatc 900
gtctttgagg acgtgcgcgg cctggagggc atcaaggagt ccacgctgca cctcacccca 960
ggccccacca gccccttcaa ggcctttgca ggcgcagacg gcaccggcat caccctcaac 1020
atcgcggtcg ccaacggcct cggcaatgcc aagaagctca tcaagcagct ggctgcaggc 1080
gagagcaagt acgacttcat cgaggtcatg gcctgccccg gcggctgcat cggcggcggc 1140
ggccagccgc gcagcgcgga caagcagatc ctgcagaagc gccaggcggc catgtacgac 1200
ctggacgagc gcgcggtgat ccggcgcagc cacgagaacc cgctgattgg cgcgctgtat 1260
gagaagttcc tgggcgagcc caacggccac aaggcgcacg agctgctgca cacgcactac 1320
gtggccggcg gcgtgcccga tgagaagtga 1350

<210> 178
<211> 1311
<212> DNA
<213> Chlorella fusca

<400> 178
atgtgttgcc ccgtggttgc aagtaggcac gcagggcgtg caaggcatgt tgctgtccgt 60
gcagcagggc caacatctga gtgtgattgt cctccaacac ctcaggccaa gctgcctcac 120
tggcagcagg ctctggatga gctcgccaag cccaaggaga gcaggaggtt gatgatcgcg 180
caaatcgcct ccgctgttcg tgtcgctatt gctgagacca ttggcttggc cccaggagat 240
gtcaccattg ggcagctcgt gactgggctg cgtatgcttg gctttgatta tgtctttgac 300
accctgtttg gtgctgacct gaccattatg gaggagggaa cggagctgct gcatcgcctg 360
caggaccatc tggagcagca ccccaacaag gaggagccac tgcccatgtt caccagttgc 420
tgcccaggct gggttgccat ggttgaaaag agcaatcctg agctcatccc ctacctgtca 480
tcttgcaagt cgcctcagat gatgcttggg gccgttatca agaactacta tgcacagcag 540
gttggagtgc agcccagtga catctgcaac gtgtcagtca tgccatgcgt acgcaagcag 600
ggagaggctg accgggagtg gttcaacacc acaggtgcag gccttgcccg tgatgttgat 660
catgtggtga ctactgctga ggttggtaag atattcctgg agcgtggcat caagctgaat 720
gagctgccag agagcaactt tgacaacccc attggcgagg gcacaggtgg tgctctgctg 780
tttggcacca ctggaggtgt catggaggca gcacttcgca cagtctatga agtggtgacc 840
cagaagccca tgggtcgtgt tgactttgag gaggtgcgag gccttgaagg aatcaaggag 900
gcagagatca cactcaagcc aggagacgac agcccattca aagccttcgc aggagctgat 960
gggcagggca tcacgctcaa gattgcagta gccaatgggc ttggcaatgc caagaagctc 1020
atcaagagcc tgtcagaggg caaggccaag tatgatttca ttgaggtcat ggcatgccct 1080
ggtggctgca ttggcggagg cggtcagccc cgcagtactg acaagcagat cctgcagaag 1140
cgccagcagg ctatgtacaa cctggatgag cgcagtacca tccgccgcag ccatgataac 1200
ccattcatcc aggcgctgta tgacaagttc ctaggcgcac ccaacagcca caaggcacat 1260

gatctgctgc acacacacta tgtggcaggt ggaattccag aggagaagtg a 1311

<210> 179
<211> 717
<212> DNA
<213> Artificial sequence

<220>
<223> Green Fluorescent Protein

<400> 179
atggccaagg gcgaggagct gttcaccggt gtggtcccca tcctggtgga gctggacggc 60

gacgtgaacg gccacaagtt ctccgtctcc ggcgagggtg agggtgacgc cacctacggc 120

aagctgaccc tgaagttcat ctgcaccacc ggcaagctgc ccgtgccctg gcccaccctg 180

gtcaccaccc tgacctacgg tgtgcagtgc ttctcccgct accccgacca catgaagcag 240

cacgacttct tcaagtccgc catgcccgag ggctacgtgc aggagcgcac catcttcttc 300

aaggacgacg gcaactacaa gacccgcgcc gaggtcaagt tcgagggcga caccctggtg 360

aaccgcatcg agctgaaggg catcgacttc aaggaggacg gcaacatcct gggccacaag 420

ctggagtaca actacaactc ccacaacgtg tacatcatgg ccgacaagca gaagaacggc 480

atcaaggtga acttcaagat ccgccacaac atcgaggacg gctccgtgca gctggccgac 540

cactaccagc agaacacccc catcggcgat ggccccgtgc tgctgcccga caaccactac 600

ctgtccatcc agtccgccct gtccaaggac cccaacgaga agcgcgacca catggtcctg 660

ctggagttcg tcaccgctgc cggcatcacc cacggcatgg acgagctgta caagtaa 717

<210> 180
<211> 320
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 180
atccgtagtt atccttatgg ccatcttagc gcagttgggt caggggctgg cgacgcgctg 60

ctgacgcgca agtgaatggc ccaacaagtc gcctcgcggt cgctgtcggc gccaaacccg 120

cagctgcatc caccagattc acttgttaga tcgacctagg ttgcgggacc ggaggcggct 180

cgctgtgcaa gcgcggtgac ctcgtacggc ggcatggatc gccatctcga ttcgcgcggc 240

agaatcgggc cccgcgcaca tttaagccgc gggcgagact catttcgtta cgtgcatcga 300

ttaacagctt ctggacctga 320

<210> 181
<211> 580
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 181
ttaaacgtcg tacgtccaag tataactaag ccgacgtcga cccactctag aggatcgatc 60

cccgctccgt gtaaatggag gcgctcgttg atctgagcct tgccccctga cgaacggcgg 120

tggatggaag atactgctct caagtgctga agcggtagct tagctccccg tttcgtgctg 180

atcagtcttt ttcaacacgt aaaaagcgga ggagttttgc aattttgttg gttgtaacga 240

tcctccgttg attttggcct ctttctccat gggcgggctg ggcgtatttg gcagttgggt 300

caggggctgg cgacgcgctg ctgacgcgca agtgaatggc ccaacaagtc gcctcgcggt 360

cgctgtcggc gccaaacccg cagctgcatc caccagattc acttgttaga tcgacctagg 420

ttgcgggacc ggaggcggct cgctgtgcaa gcgcggtgac ctcgtacggc ggcatggatc 480

gccatctcga ttcgcgcggc agaatcgggc cccgcgcaca tttaagccgc gggcgagact 540

catttcgtta aatctgatac atgctattca gatcttacaa 580

<210> 182
<211> 580
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 182
tcttccatcg taaatctagc atcgattagc ccgacgtcga cccactctag aggatcgatc 60

cccgctccgt gtaaatggag gcgctcgttg atctgagcct tgccccctga cgaacggcgg 120

tggatggaag atactgctct caagtgctga agcggtagct tagctccccg tttcgtgctg 180

atcagtcttt ttcaacacgt aaaaagcgga ggagttttgc aattttgttg gttgtaacga 240

tcctccgttg attttggcct ctttctccat gggcgggctg ggcgtatttg gcagttgggt 300

caggggctgg cgacgcgctg ctgacgcgca agtgaatggc ccaacaagtc gcctcgcggt 360

cgctgtcggc gccaaacccg cagctgcatc caccagattc acttgttaga tcgacctagg 420

ttgcgggacc ggaggcggct cgctgtgcaa gcgcggtgac ctcgtacggc ggcatggatc 480

gccatctcga ttcgcgcggc agaatcgggc cccgcgcaca tttaagccgc gggcgagact 540

catttcgtta atctgtaata atctagtcga ggcattcaag 580

<210> 183
<211> 777
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 183
atctgtaata atctagtcga ggcattcaag atggccaagg gcgaggagct gttcaccggt 60

gtggtcccca tcctggtgga gctggacggc gacgtgaacg gccacaagtt ctccgtctcc 120

ggcgagggtg agggtgacgc cacctacggc aagctgaccc tgaagttcat ctgcaccacc 180

ggcaagctgc ccgtgccctg gcccaccctg gtcaccaccc tgacctacgg tgtgcagtgc 240

ttctcccgct accccgacca catgaagcag cacgacttct tcaagtccgc catgcccgag 300

ggctacgtgc aggagcgcac catcttcttc aaggacgacg gcaactacaa gacccgcgcc 360

gaggtcaagt tcgagggcga caccctggtg aaccgcatcg agctgaaggg catcgacttc 420

aaggaggacg gcaacatcct gggccacaag ctggagtaca actacaactc ccacaacgtg 480

tacatcatgg ccgacaagca gaagaacggc atcaaggtga acttcaagat ccgccacaac 540

atcgaggacg gctccgtgca gctggccgac cactaccagc agaacacccc catcggcgat 600

ggccccgtgc tgctgcccga caaccactac ctgtccatcc agtccgccct gtccaaggac 660

dccaacgaga agcgcgacca catggtcctg ctggagttcg tcaccgctgc cggcatcacc 720

cacggcatgg acgagctgta caagtaaaac tggcttaaat cgttaacaat cgtgtga 777

<210> 184
<211> 320
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic sequence

<400> 184
aactggctta aatcgttaac aatcgtgtga ccgacgtcga cccactctag aggatcgatc 60

cccgctccgt gtaaatggag gcgctcgttg atctgagcct tgccccctga cgaacggcgg 120

tggatggaag atactgctct caagtgctga agcggtagct tagctccccg tttcgtgctg 180

atcagtcttt ttcaacacgt aaaaagcgga ggagttttgc aattttgttg gttgtaacga 240

tcctccgttg attttggcct ctttctccat gggcgggctg ggcgtatttg gatttaacat 300

aactgtcgat taccgtgcga 320

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
11 August 2005 (11.08.2005)

PCT

(10) International Publication Number
WO 2005/072262 A3

(51) International Patent Classification:
C12P 3/00 (2006.01) *C12N 1/13* (2006.01)
C12N 1/00 (2006.01)

(21) International Application Number:
PCT/US2005/001983

(22) International Filing Date: 21 January 2005 (21.01.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/763,712 21 January 2004 (21.01.2004) US

(71) Applicant *(for all designated States except US)*: **SOLAZYME, INC.** [US/US]; Edison Technology Park, 3475-T Edison Way, Menlo Park, CA 94025 (US).

(72) Inventor: **DILLON, Harrison, F.**; Solazyme, Inc., Edison Technology Park, 3475-T Edison Way, Menlo Park, CA 94025 (US).

(74) Agent: **HALIDAY, Emily M.**; QUINE INTELLECTUAL PROPERTY, LAW GROUP, P.C., P.O. Box 459, Alameda, CA 94501 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

(88) Date of publication of the international search report:
2 November 2006

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHODS AND COMPOSITIONS FOR EVOLVING MICROBIAL HYDROGEN PRODUCTION

(57) Abstract: The invention provides methods and compositions for engineering cells to generate large amounts of hydrogen. Genes that are involved in hydrogen production pathways and genes that are upregulated when cells are exposed to conditions conducive to the generation of hydrogen are mutagenized according to disclosed protocols. Microbes containing nucleic acid constructs are screened or selected for the ability to generate an increased amount of hydrogen. Methods of producing hydrogen are also disclosed.

WO 2005/072262 A3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US05/01983

A. CLASSIFICATION OF SUBJECT MATTER
IPC: **C12P 3/00**(2006.01);**C12N 1/00**(2006.01),**1/13**(2006.01)

USPC: 435/168,257.2,257.6
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 435/168, 257.2, 257.6

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	GHIRARDI, M. ET AL. Algal Hydrogen Photoproduction, Program Review Meeting, Berkeley, CA, May 2003, http://www1.eere.energy.gov/hydrogenandfuelcells/pdfs/merit03/42_nrel_maria_ghirardi.pdf	1
X	US Patent App. Pub. No: 2003/0162273 A1 (MELIS ET AL) 28 August 2003 (28.08.2003), see entire document, especially page 1, paragraph 8; page 2, paragraph 17; and page 24, claims 1-2, 4 and 7	1-2, 16-17 and 21

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
21 July 2006 (21.07.2006)	25 AUG 2006
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450 Alexandria, Virginia 22313-1450 Facsimile No. (571) 273-3201	Authorized officer Walter Schlapkohl Telephone No. (571) 272-0500

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US05/01983

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of any additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-3, 13-28 (ferredoxin, SEQ ID NO: 172)

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.